

STIC-Biotech/ChemLib

79093

From: Portner, Ginny  
Sent: Wednesday, October 30, 2002 1:58 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/910,186  
Importance: High

Please search SEQ ID Nos 9 and 10, please back translate SEQ ID NO 10. Thanks.

Ginny Portner  
CM1, Art Unit 1645  
Room 7e13  
Mail box 7e12  
(703) 308-7543

Edward Hart  
Technical Info. Specialist  
STIC/Biotech  
CMI 6B02 Tel: 305-9203

RECEIVED  
OCT 30 2002  
STIC

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 11/4/02  
Date Completed: 11/8/02  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: 141  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: 10/6  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rapn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

*The Pending database search results should not be left in the case because they contain data that is confidential.*

According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:

**Published\_Applications\_NA** contains nucleic acid sequences; the search results will have the extension **.rnpb**.

**Published\_Applications\_AA** contains amino acid sequences; the search results will have the extension **.rapb**.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicants submit changes to the application, they may request that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

**Sequences in the PGPub database are public information; it is permissible to leave these results in the case.**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 13:45:26 ; Search time 55 Seconds  
(without alignments)  
922.504 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MIPFNFYSYTNNSLLKDI.....NYASLLESTSHWGFVPVSE 450

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_101002.\*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
  - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
  - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
  - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
  - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
  - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
  - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
  - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
  - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
  - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
  - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
  - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
  - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
  - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
  - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
  - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
  - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
  - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
  - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
  - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
  - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
  - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
  - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2391	100.0	450	21	AAV77136
2	2391	100.0	450	22	AAV77136
3	2386	99.8	1291	20	AAV05814
4	2317	96.9	462	19	AAW68397
5	1599	66.5	399	21	AAV78982
6	867.5	35.3	451	22	AAV04093
7	801.5	33.5	451	19	AAW68398
8	589.5	24.7	449	21	AAV77137
9	589.5	24.7	449	22	AAV04094
10	586.5	24.5	837	21	AAV77140

11	586.5	24.5	1067	21	AAV93307	A manganese supero
12	586.5	24.5	1092	21	AAV93310	A manganese supero
13	586.5	24.5	1296	17	AAV93308	C. botulinum type
14	581	24.3	1070	21	AAV93311	A manganese supero
15	581	24.3	1095	21	AAV93311	A manganese supero
16	581	24.3	1291	19	AAW68392	Clostridium botuli
17	578.5	24.2	452	19	AAW68396	Clostridium botuli
18	577.5	24.2	1295	23	AAU99339	Clostridium botuli
19	575.5	24.1	451	19	AAW68395	Synthetic botulinu
20	575.5	24.1	449	21	AAV77139	Botulism toxin hea
21	575.5	24.1	449	22	AAV04167	Clostridium botuli
22	575.5	24.1	473	19	AAW68400	Botulism toxin hea
23	573	24.0	848	22	AAV04082	Botulism toxin hea
24	568.5	23.8	847	22	AAV04081	Botulism toxin hea
25	565	23.6	472	19	AAW68393	Clostridium botuli
26	564	23.6	1059	21	AAV93309	A manganese supero
27	564	23.6	1084	21	AAV93312	A manganese supero
28	558	23.3	1092	22	AAE07900	C. botulinum C2 tr
29	557.5	23.3	432	22	AAV77142	Native botulinum n
30	557.5	23.3	437	22	AAV04088	Botulism toxin hea
31	557.5	23.3	438	17	AAV95008	Type A neurotoxin
32	557.5	23.3	438	19	AAW68389	Clostridium botuli
33	557.5	23.3	438	21	AAV77134	Synthetic botulinu
34	557.5	23.3	445	19	AAW68391	Clostridium botuli
35	557.5	23.3	462	17	AAV95009	Type A neurotoxin
36	557.5	23.3	462	19	AAW68390	Clostridium botuli
37	556.5	23.3	434	22	AAV04089	Botulism toxin hea
38	556.5	23.3	435	22	AAV04090	Botulism toxin hea
39	553.5	23.1	419	22	AAV04095	Botulism toxin hea
40	551	23.0	1032	22	AAE07901	C. botulinum C2 tr
41	550	23.0	472	19	AAW68394	Clostridium botuli
42	548	22.9	685	22	AAV07893	Modified clostridi
43	546	22.8	431	18	AAW09014	Immunogenic type F
44	546	22.8	432	22	AAV04096	Botulism toxin hea
45	546	22.8	432	22	AAV04103	Botulism toxin hea

## ALIGNMENTS

### RESULT 1

AAV77136  
ID AAV77136 standard; Protein: 450 AA.

AC AAV77136;

XX 08-MAY-2000 (first entry)

DT Synthetic botulinum neurotoxin serotype C (BoNTC) C-terminal fragment.

DE Botulinum neurotoxin; heavy chain; BoNT; serotype C;

XX C-terminal fragment; Venezuelan equine encephalitis virus replicon;

KW VEE; botulism; vaccine; diagnosis; drug screening.

XX Clostridium botulinum.

OS Synthetic.

PN WO200002524-A2.

XX 20-JAN-2000.

PF 09-JUL-1999; 99WO-US15570.

PR 10-JUL-1998; 98US-0092416.

XX 12-MAY-1999; 99US-0133870.

XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.

PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;

DR WPI: 2000-160827/14.

XX N-PSDB; AA287214.



PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
PT toxin serotypes A-G, is used for inducing an immune response against  
PT botulinum -  
XX  
PS Claim 24; Page 41-42; 54pp; English.  
XX  
CC The invention relates to novel vaccines that induce a protective immune  
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
CC DNA construct comprising a vector, and at least one nucleic acid  
CC fragment comprising a C-terminal heavy chain fragment (HC) from BoNT  
CC serotypes A-G. In preferred embodiments of the invention, the vector is  
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
CC this vector results in the production of large amounts of a protein  
CC encoded by a sequence cloned into the replicon. The constructs are used  
CC to produce vaccines against botulinum. The proteins can also be used as  
CC diagnostic tools for the diagnosis of botulinum. The transformed host  
CC cells can be used to analyse the effectiveness of drugs and agents which  
CC inhibit toxin effects. The vaccine currently used against botulinum is  
CC dangerous and expensive to produce, and contains formalin, which is very  
CC painful for the recipient. Also, the vaccine is incomplete, in that only  
CC 5 of the 7 serotypes are represented in the formulation. The novel  
CC vaccine of overcomes these problems, as it is easily purified, and  
CC available in large quantities. It is also expressed in the lymph nodes  
CC for a better immune response. Sequences AAY7134-Y7139 represent  
CC synthetic BoNT Hc fragments used in the present invention. The DNA  
CC encoding these sequences had been optimised for codon usage for  
CC expression in yeast.  
XX  
SQ Sequence 450 AA;  
Query Match 100.0%; Score 2391; DB 21; Length 450;  
Best Local Similarity 100.0%; Pred. No. 9.6e-164;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTIPNIFSYNNLLKDIINEYFNNDKSKLSLQNRKNTLVDTSGYNAEVEEGDVQL 60  
Db 1 MTIPNIFSYNNLLKDIINEYFNNDKSKLSLQNRKNTLVDTSGYNAEVEEGDVQL 60  
QY 61 NPIFFDFKLSSGDEDRGKVIQTQENIVNMYESFSISFWIRINKWVSNLPGYTIIDS 120  
Db 61 NPIFFDFKLSSGDEDRGKVIQTQENIVNMYESFSISFWIRINKWVSNLPGYTIIDS 120  
QY 121 VKNNGWSIGIISNPLVFTLKQNEDESGSINFSYDIGNNAPGYNKWFVVTNNMGNMK 180  
Db 121 VKNNGWSIGIISNPLVFTLKQNEDESGSINFSYDIGNNAPGYNKWFVVTNNMGNMK 180  
QY 181 IYNGKLIDTIKVKELTGINFSTKITEPINKIPTGLITSDSNINMWIRDFYIFAKELD 240  
Db 181 IYNGKLIDTIKVKELTGINFSTKITEPINKIPTGLITSDSNINMWIRDFYIFAKELD 240  
QY 241 GKDINTLNSLOYTNVVDYGNLRYNKEYYVNDYLNRYMYANSRQVFNTRNNND 300  
Db 241 GKDINTLNSLOYTNVVDYGNLRYNKEYYVNDYLNRYMYANSRQVFNTRNNND 300  
QY 301 FNEGYKIIIRKIRGNNTDTRVGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAI 360  
Db 301 FNEGYKIIIRKIRGNNTDTRVGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIYAI 360  
QY 361 GLREQTKDINDNIIFQIOPMNTYYASQIFKSNFNGENTSGICISGYFRLLGGDWYRH 420  
Db 361 GLREQTKDINDNIIFQIOPMNTYYASQIFKSNFNGENTSGICISGYFRLLGGDWYRH 420  
QY 421 NVLPTVKOGNYASLLESTSTHWGFPVYSE 450  
Db 421 NVLPTVKOGNYASLLESTSTHWGFPVYSE 450  
RESULT 2  
AAB04092  
ID AAB04092 standard; Protein; 450 AA.  
XX  
AC AAB04092;

XX 11-APR-2001 (first entry)  
XX Botulinum toxin heavy chain C-terminal sequence (serotype C).  
XX  
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;  
KW recombinant vector; antigen; immune response; vaccine; bacterium;  
KW infection.  
XX  
OS Synthetic.  
OS Clostridium botulinum.  
XX  
PN W0200067700-A2.  
XX  
PD 16-NOV-2000.  
XX  
PF 12-MAY-2000; 2000WO-US12890.  
XX  
PR 12-MAY-1999; 99US-0133865.  
PR 12-MAY-1999; 99US-0133866.  
PR 12-MAY-1999; 99US-0133867.  
PR 12-MAY-1999; 99US-0133868.  
PR 12-MAY-1999; 99US-0133869.  
PR 12-MAY-1999; 99US-0133873.  
PR 29-JUL-1999; 99US-0146192.  
XX  
FA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
XX  
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
XX  
XX WPI; 2001-016048/02.  
XX N-PSDB; AAA54486.  
XX  
XX New nucleic acids encoding the carboxy- or amino-terminal portions of  
XX the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
XX vaccine against botulism  
XX  
XX Claim 3; Fig 5b; 73pp; English.  
XX  
XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
XX chain and then posttranslationally nicked, forming a dichain  
XX consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
XX remain linked by a disulfide bond. Nucleic acids encoding the  
XX carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
XX chain of botulinum neurotoxin (BoNT) can be used in recombinant  
XX expression vectors and expressed in transformed cells to produce  
XX peptide antigens useful for eliciting an immune response to give  
XX protective immunity against botulinum neurotoxin, which causes  
XX botulism. The nucleic acids are expressible in a recombinant  
XX organisms such as Escherichia coli or Pichia pastoris. The use  
XX of recombinant nucleic acids are advantageous since it eliminates  
XX the need to culture large quantities of hazardous toxin-producing  
XX bacterium. Production yield from the genetically engineered product  
XX is also high and cost of production is lower. The nucleic acids can  
XX be derived from Clostridium botulinum serotypes A-G.  
XX  
SQ Sequence 450 AA;  
Query Match 100.0%; Score 2391; DB 22; Length 450;  
Best Local Similarity 100.0%; Pred. No. 9.6e-164;  
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTIPNIFSYNNLLKDIINEYFNNDKSKLSLQNRKNTLVDTSGYNAEVEEGDVQL 60  
Db 1 MTIPNIFSYNNLLKDIINEYFNNDKSKLSLQNRKNTLVDTSGYNAEVEEGDVQL 60  
QY 61 NPIFFDFKLSSGDEDRGKVIQTQENIVNMYESFSISFWIRINKWVSNLPGYTIIDS 120  
Db 61 NPIFFDFKLSSGDEDRGKVIQTQENIVNMYESFSISFWIRINKWVSNLPGYTIIDS 120  
QY 121 VKNNGWSIGIISNPLVFTLKQNEDESGSINFSYDIGNNAPGYNKWFVVTNNMGNMK 180  
Db 121 VKNNGWSIGIISNPLVFTLKQNEDESGSINFSYDIGNNAPGYNKWFVVTNNMGNMK 180



XX (OPHI-) OPHIDIAN PHARM INC.

XX Thalley BS, Williams JA;

XX WPI; 1998-230234/20.

DR N-PSDB; AAV30588.

XX Host cell containing recombinant expression vector encoding

PT Clostridium botulinum type B or E toxin - useful to treat humans

PT and other animals at risk of intoxication with clostridial toxin

XX Example 45; Page 339-341; 428pp; English.

XX This is the amino acid sequence of the histidine-tagged C fragment  
CC of Clostridium botulinum (Stockholm strain) type C1 neurotoxin,  
CC encoded by a DNA sequence (see AAV30588) in plasmid pTH18b. This  
CC vector is used to express B0C soluble C fragment in Escherichia  
CC coli host cells, and the recombinant C fragment was purified on an  
CC affinity column. The invention relates to recombinant proteins  
CC derived from C. botulinum toxins, especially type B and type E  
CC toxins. Methods are provided which allow for the isolation of  
CC soluble recombinant proteins free of significant endotoxin  
CC contamination. Preferred hosts for production of recombinant  
CC proteins are E. coli, insect cells and yeast cells. The  
CC recombinant toxins are used as immunogens for the production of  
CC vaccines and antitoxins that are useful in the treatment of humans  
CC and animals at risk of intoxication with clostridial toxin.

XX Sequence 462 AA;

Query Match 96.98; Score 2317; DB 19; Length 462;  
Best Local Similarity 99.88; Pred. No. 2e-158;  
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLIKDIINEYNNINDSKILSLQNKNTLVDTSGYNAEVEEGDVQLNPIFPDFKLGS 73

Db 26 ALLKDIINEYNNINDSKILSLQNKNTLVDTSGYNAEVEEGDVQLNPIFPDFKLGS 85

QY 74 GDRGKVIQTQENIVNMYTESFSISFWIRINKWVNLPGYTIIDSVKNSGWSIGIS 133

Db 86 GDRGKVIQTQENIVNMYTESFSISFWIRINKWVNLPGYTIIDSVKNSGWSIGIS 145

QY 134 NFLVFTLKQNEDESGINFSYDIENNAPGVNKPFVTVTNMGNMKIYINGKLIDTIKV 193

Db 146 NFLVFTLKQNEDESGINFSYDIENNAPGVNKPFVTVTNMGNMKIYINGKLIDTIKV 205

QY 194 KELTGINFSKTITPEINKIPDTGLITSDSDINNMWIRDFYIFAKELGDKDINILFNSLO 253

Db 206 KELTGINFSKTITPEINKIPDTGLITSDSDINNMWIRDFYIFAKELGDKDINILFNSLO 265

QY 254 TTVVVDYWGNDLRYNKEYMYNIDYLNRYMYANSQIIVENTRRNNDFNKGKIIIKRIR 313

Db 266 TTVVVDYWGNDLRYNKEYMYNIDYLNRYMYANSQIIVENTRRNNDFNKGKIIIKRIR 325

QY 314 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDYIAGLREQTKDINDNI 373

Db 326 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDYIAGLREQTKDINDNI 385

QY 374 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYFRGLGGDWYRNYLVPVYKQNYA 433

Db 386 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYFRGLGGDWYRNYLVPVYKQNYA 445

QY 434 SLESTSTHWGFPVPSVE 450

Db 446 SLESTSTHWGFPVPSVE 462

RESULT 5

AAV78982

ID AAV78982 standard; Protein; 399 AA.

XX

AC AAV78982;

XX 20-JUN-2000 (first entry)

XX C. botulinum type D toxin amino acid sequence.

XX Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;  
KW protect.

XX Clostridium botulinum.

XX W0200005252-A1.

XX 03-FEB-2000.

XX 20-JUL-1999; 99WO-IB01301.

XX 22-JUL-1998; 98ZA-0006538.

XX (AGRI-) AGRIC RES COUNCIL.

XX De Bruyn EE, Botha AD;

XX WPI; 2000-205375/18.

XX N-PSDB; AA298630, AA298631.

XX Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin  
PT type D, useful in vaccines for protection against botulism, comprises  
PT at least one amino acid mutation not present in the wild type D  
PT neurotoxins -

XX Claim 3, 4; Page 54-57; 66pp; English.

XX This sequence represents the amino acid sequence of a synthetic  
CC non-toxic immunogenic derivative of Clostridium botulinum type D toxin  
CC (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle and  
CC sheep, and usually results in the death of the affected or poisoned  
CC animal. The non-toxic immunogenic fragments of the C. botulinum  
CC neurotoxin are useful in vaccines to protect animals (e.g. humans,  
CC cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic  
CC fragments can be produced relatively simply and inexpensively  
CC (specifically by fermentation techniques). As the fragments are not  
CC toxic the risk to production staff is reduced.

XX Sequence 399 AA;

Query Match 66.5%; Score 1589; DB 21; Length 399;  
Best Local Similarity 76.08; Pred. No. 3.3e-106;  
Matches 307; Conservative 33; Mismatches 56; Indels 8; Gaps 4;

QY 50 AEVSEEGDVQLNPIFPDFKLGSSEDRGKVIYVQENIVNMYTESFSISFWIRINKWV 109

Db 1 AEVREGVQLNPIFPDFKLGSSEDRGKVIYVQENIVNMYTESFSISFWIRINKWV 60

QY 110 SNLPGYTIIDSVKNSGWSIGIISNLFVTLKQNEDESGINFSYDIENNAPGVNKWFFV 169

Db 61 SNLPGYTIIDSVKNSGWSIGIISNLFVTLKQNEDESGINFSYDIENNAPGVNKWFFV 120

QY 170 TVTNMGNMKIYINGKLIDTIKYKELTGINFSKTITPEINKIPDTGLITSDSDINMWI 229

Db 121 ITTNMGNMKIYINGKLIDTIKYKELTGINFSKTITPEINKIPDTGLITSDSDINMWI 180

QY 230 RDFYIFAKELGDKDINILFNSLOTVVVDYWGNDLRYNKEYMYNIDYLNRYMYANSRQ 289

Db 181 RDFYIFAKELGDKDINILFNSLOTVVVDYWGNDLRYNKEYMYNIDYLNRYMYANSRQ 240

QY 290 IVFNTRRNNDFNKGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYA 349

Db 241 IVFNTRRNNDFNKGKIIIKRIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYA 295

QY 350 DNHST-EDIYAGLREQTKD-INDNIIFQIQPMNNTYYASQIFKSNFNGENISGICISG 407

Db 296 PSRNLGTDLVPLGALDQPMDEIRKYSFIQPCNTFDYASQLFSLSSNATNRLGILSIG 355







PT botulinum -  
PS Example 3; Page 49; 54pp; English.  
XX  
XX The invention relates to novel vaccines that induce a protective immune  
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
CC and G (BoNTA-BONGT). The vaccine of the invention is novel recombinant  
CC DNA construct comprising a vector and at least one nucleic acid  
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT  
CC serotypes A-G. In preferred embodiments of the invention, the vector is a  
CC Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this  
CC vector results in the production of large amounts of a protein encoded by  
CC a sequence cloned into the replicon. The constructs are used to produce  
CC vaccines against botulinum. The proteins can also be used as diagnostic  
CC tools for the diagnosis of botulinum. The transformed host cells can be  
CC used to analyse the effectiveness of drugs and agents which inhibit toxin  
CC effects. The vaccine currently used against botulinum is dangerous  
CC and expensive to produce, and contains formalin, which is very painful  
CC for the recipient. Also, the vaccine is incomplete, in that only 5 of  
CC the 7 serotypes are represented in the formulation. The novel vaccine  
CC of overcomes these problems, as it is easily purified, and available in  
CC large quantities. It is also expressed in the lymph nodes for a better  
CC immune response. The present sequence represents the native BoNTA heavy  
CC chain used in an exemplification of the present invention.  
XX  
SQ Sequence 837 AA;  
Query Match 24.5%; Score 586.5; DB 21; Length 837;  
Best Local Similarity 31.6%; Pred. No. 8.2e-34;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;  
QY 3 IPFNIFSYTNNSLLKDIINEYFNNDKSKILSNQKNTLVDTSGYNAEVSSEGDVQLNP 62  
DB 396 IPFOLSKYVDNQRLLSFTTEYIKNIINTSLNRYESNHLIDLSRYASKINIGSKVNFDP 455  
QY 63 IPFPD---FKLGSGEDRGKVIYTONENIYNSWYFSFSFWIRNKWVSNLP---GYT 116  
DB 456 IDKQIQIOLFNLESS-----KIEVLKNAIYNSMYENFSTFWIRPKYFNSISLNNEYT 510  
QY 117 IDSVKYNNSGWSIGIISNLFVTLKQNEDESEQINSYDISNNAPGY-NKWEFVTVTNM 175  
DB 511 IINCENNSGKWSLNGEIIWTLQDTQEIQRVFKYSQMINISDYINRWIFVTINNR 570  
QY 176 MGNKIYINGKLIDTIKVKELTGINFESKTTTFEINKIPDPGLITSDSDINNMWIRDFYIF 235  
DB 571 LNSKIYINGRLIDQKPTISLGNTHASNIMFKLDGCRD-----HRYIWKIFNLF 622  
QY 236 AKELDGKDINILFNSLOQYTNVVDYWGNDLRYNKKEYYMNIDYLNRYMYANS----- 287  
DB 623 DKELNEKEIKLDYDNQNSGILKDFWGDYLDQKPYMLNLYDPNKYVDVNVYNGIRGYMY 682  
QY 288 ----RQIVFNTR-RNNNDFNEGYKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYLEM 342  
DB 683 LKGPGRGYMTNIVNLSLYRGTFEIKKYASGNKDNIVRNDRVYINVVVKNKEURL-- 740  
QY 343 KNETMYADNHSTEDYATGLREQDKINDIMLIFQIPMNNTYIYASQIFKSNPENGNISG 402  
DB 741 -----ATNAQAGVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGND 793  
QY 403 ICSITVFRIGGDYRNYNLYPVTVKQNTASLEST-----THWGFVPVSE 450  
DB 794 IGFIFGHQF-----NNIAKLVASWNYNRQIERSRSLTLCGSWEFIPVDD 836  
RESULT 11  
AA93307  
ID AA93307 standard; protein; 1067 AA.  
XX  
XX  
AC AA93307;  
XX  
XX  
DT 04-SEP-2000 (first entry)  
XX  
XX A manganese superoxide dismutase (Mn-SOD) construct.

XX Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
KW neuronal cell targeting component; NCTC; neuronal disease;  
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
KW Huntington's disease; motor neuron disease;  
KW botulinum neurotoxin serotype A.  
XX  
OS Synthetic.  
OS Bacillus steatothermophilus.  
OS Clostridium botulinum.  
XX  
PN W0200028041-A1.  
XX  
PD 18-MAY-2000.  
XX  
PF 05-NOV-1999; 99WO-GB03699.  
XX  
PR 05-NOV-1998; 98GB-0024282.  
XX  
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
XX  
PI Shone CC, Sutton JM, Hallis B, Silman N;  
XX WPI; 2000-376553/32.  
DR  
XX  
XX Novel composition, comprising superoxide dismutase linked by a  
PT cleavable linker to a neuronal cell targeting component useful for  
PT delivering superoxide dismutase to neuronal cells to treat ischemia -  
PS Disclosure; Page 41-44; 65pp; English.  
XX  
CC The present sequence represents a construct of the invention, comprising  
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that  
CC can be cleaved by thrombin, and a heavy chain derived from botulinum  
CC neurotoxin serotype A. The specification describes a composition for  
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,  
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).  
CC This component has a domain that binds to a neuronal cell and a  
CC domain that translocates the SOD of the composition into the neuronal  
CC cell. After translocation, the linker is cleaved to release the SOD.  
CC The composition is useful for treating neuronal diseases caused or  
CC augmented by oxidative stress, such as ischemic stroke, trauma,  
CC Parkinson's disease, Huntington's disease and motor neuron diseases.  
XX  
SQ Sequence 1067 AA;  
Query Match 24.5%; Score 586.5; DB 21; Length 1067;  
Best Local Similarity 31.8%; Pred. No. 1.1e-33;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;  
QY 3 IPFNIFSYTNNSLLKDIINEYFNNDKSKILSNQKNTLVDTSGYNAEVSSEGDVQLNP 62  
DB 620 IPFOLSKYVDNQRLLSFTTEYIKNIINTSLNRYESNHLIDLSRYASKINIGSKVNFDP 679  
QY 63 IPFPD---FKLGSGEDRGKVIYTONENIYNSWYFSFSFWIRNKWVSNLP---GYT 116  
DB 680 IDKQIQIOLFNLESS-----KIEVLKNAIYNSMYENFSTFWIRPKYFNSISLNNEYT 734  
QY 117 IDSVKYNNSGWSIGIISNLFVTLKQNEDESEQINSYDISNNAPGY-NKWEFVTVTNM 175  
DB 735 IINCENNSGKWSLNGEIIWTLQDTQEIQRVFKYSQMINISDYINRWIFVTINNR 794  
QY 176 MGNKIYINGKLIDTIKVKELTGINFESKTTTFEINKIPDPGLITSDSDINNMWIRDFYIF 235  
DB 795 LNSKIYINGRLIDQKPTISLGNTHASNIMFKLDGCRD-----HRYIWKIFNLF 846  
QY 236 AKELDGKDINILFNSLOQYTNVVDYWGNDLRYNKKEYYMNIDYLNRYMYANS----- 287  
DB 847 DKELNEKEIKLDYDNQNSGILKDFWGDYLDQKPYMLNLYDPNKYVDVNVYNGIRGYMY 906  
QY 288 ----RQIVFNTR-RNNNDFNEGYKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYLEM 342  
DB 907 LKGPGRGYMTNIVNLSLYRGTFEIKKYASGNKDNIVRNDRVYINVVVKNKEURL-- 964

Qy 343 KNETMYADNHSTEDIYAIGUREOTKDINDNIIFQIQPMNTYYIASQIFKSNFNGENISG 402  
 Db 965 -----ATNASQAGVERKILSALTEIPDVGNSLQVVMKSKNDQGITNKC-KMNLQNNNGND 1017  
 Qy 403 ICSIGTYRFLRGDWRHNYLPTVTKOQNYASLLESTG-----THWGFVPVSE 450  
 Db 1018 IGFIFGHQF-----NNIAKLVASNWTYNROIERSRSLGCSWEFIPVDD 1060

RESULT 12  
 AAY93310  
 ID AAY93310 standard; protein; 1092 AA.  
 AC AAY93310;  
 XX  
 DT 04-SEP-2000 (first entry)  
 XX  
 DE A manganese superoxide dismutase (Mn-SOD) construct.  
 XX  
 KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
 KW neuronal cell targeting component; NCIC; neuronal disease;  
 KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
 KW Huntington's disease; motor neurone disease;  
 KW botulinum neurotoxin serotype A.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 OS Bacillus stearothermophilus.  
 OS Clostridium botulinum.  
 XX  
 PN WO200028041-A1.  
 PD 18-MAY-2000.  
 XX  
 PF 05-NOV-1999; 99WO-GB03699.  
 XX  
 PR 05-NOV-1998; 98GB-0024282.  
 XX  
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 XX  
 PI Shone CC, Sutton JM, Hallis B, Silman N;  
 XX WPI; 2000-376553/32.  
 DR  
 PT Novel composition, comprising superoxide dismutase linked by a  
 PT cleavable linker to a neuronal cell targeting component useful for  
 PT delivering superoxide dismutase to neuronal cells to treat ischemia -  
 PS Disclosure; Page 51-54; 65pp; English.  
 XX  
 CC The present sequence represents a construct of the invention, comprising  
 CC a mitochondrial leader sequence from human manganese superoxide  
 CC dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker  
 CC that can be cleaved by thrombin, and a heavy chain derived from  
 CC botulinum neurotoxin serotype A. The specification describes a  
 CC composition for delivery of SOD to neuronal cells. The composition  
 CC comprises SOD linked, by a cleavable linker, to a neuronal cell  
 CC targeting component (NCIC). This component has a domain that binds  
 CC to a neuronal cell and a domain that translocates the SOD of the  
 CC composition into the neuronal cell. After translocation, the linker  
 CC is cleaved to release the SOD. The composition is useful for treating  
 CC neuronal diseases caused or augmented by oxidative stress, such as  
 CC ischemic stroke, trauma, Parkinson's disease, Huntington's disease and  
 CC motor neurone diseases.  
 XX  
 SQ sequence 1092 AA;  
 Query Match 24.5%; Score 586.5; DB 21; Length 1092;  
 Best Local Similarity 31.6%; Pred. No. 1 ie-33;  
 Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPFNIFSTNNSLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEVSEGDVQLNP 62

Db 645 IPFOLSKYVDNQALLSTFTTEYIKNIINTSILNRYESNHLIDLSRYASKINIGSKVNDP 704  
 Qy 63 IFPPD---FKLSSGEDRGKVIYVTQENIYVNSMYESFISFWIRINKVWSNLP---GYT 116  
 Db 705 IDKNOIQLEFNLESS---KIEVLKNAIVNSMYENFSTFWIRIPKVFNSISLNNEYT 759  
 Qy 117 IIDSVKNSGWSIGTISNLFVFLTKQNEDESEQSINFSYDISNAPGY-NKWEFFVTNNM 175  
 Db 760 IINCMMNSGKVSLSNGEIIWLTQDTEIKQKVYKYSQMINISDYINRWIFVITNNR 819  
 Qy 176 MGNMKIYINGKLIDITIKVKELTGINFSKTIITFEINKIPDTGLTSDSDINNMWIRDFYIF 235  
 Db 820 LNSKIYINGRLIDQKPIISNLGNHASNNIMFKLDCRDT-----HRYTIWKYFNLF 871  
 Qy 236 AKELDGKDINILFNLSQYTNVVDYWGNDLRYNKYKMYNIDYLNRYMYANS----- 287  
 Db 872 DKELNEKEIKDLYDNQSNIGILKDFWGDYQYDKPYMLNLYDPNKYVDVNNVGIRGYM 931  
 Qy 288 ---RQIVFNTR-RNNNDFNEGKIIKRIKRGNTNDRVRGGDILYFDMTINKKAYNLFM 342  
 Db 932 LKGRGSMVTNIYLSLSLYRGKFIKKYASGNKDNIVRNNDRVYINVVKNKEYRL-- 989  
 Qy 343 KNETMYADNHSTEDIYAIGUREOTKDINDNIIFQIQPMNTYYIASQIFKSNFNGENISG 402  
 Db 990 -----ATNASQAGVERKILSALTEIPDVGNSLQVVMKSKNDQGITNKC-KMNLQNNNGND 1042  
 Qy 403 ICSIGTYRFLRGDWRHNYLPTVTKOQNYASLLESTG-----THWGFVPVSE 450  
 Db 1043 IGFIFGHQF-----NNIAKLVASNWTYNROIERSRSLGCSWEFIPVDD 1085

RESULT 13  
 AAR95010  
 ID AAR95010 standard; Protein; 1296 AA.  
 AC AAR95010;  
 DT 09-JUL-1996 (first entry)  
 XX  
 DE C. botulinum type A neurotoxin.  
 XX  
 KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen.  
 XX Clostridium botulinum.  
 OS  
 PN WO9612802-A1.  
 XX  
 PD 02-MAY-1996.  
 XX  
 PF 23-OCT-1995; 95WO-US13737.  
 XX  
 PR 07-JUN-1995; 95US-0480604.  
 PR 24-OCT-1994; 94US-0329154.  
 PR 16-MAR-1995; 95US-0405496.  
 PR 14-APR-1995; 95US-0422711.  
 XX  
 PA (OPHI-) OPHIDIAN PHARM INC.  
 XX  
 PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;  
 PI Williams JA;  
 XX  
 DR WPI; 1996-230603/23.  
 DR N-PSDB; AAT29244.  
 XX  
 PT Fusion proteins comprising non-toxin protein and part of toxin -  
 PT useful to form anti-toxins against Clostridium botulinum type A, and  
 PT C. difficile type toxins, and to treat C. difficile intoxication,  
 PT partic. diarrhoea  
 XX  
 PS Claim 4; Page 344-350; 434pp; English.  
 XX Clostridium botulinum type A neurotoxin (AAR95010) is processed to form



CC a dimer composed of a light and a heavy chain. It is the product of  
 CC the type A neurotoxin gene (AAR29244). The 50 kDa C-terminal portion  
 CC of the heavy chain, or C fragment (see also AAR95008), was produced  
 CC using a synthetic gene (AAR29245) having codon usage altered to improve  
 CC expression in *Escherichia coli*. Fusion proteins of the type A toxin  
 CC or C fragment, with e.g. maltose binding protein or polyhistidine  
 CC affinity tag (see also AAR95008), are used to generate neutralising  
 CC antitoxins and in vaccine compsns.

XX Sequence 1296 AA;

Query Match 24.5%; Score 586.5; DB 17; Length 1296;  
 Best Local Similarity 31.6%; Pred. No. 1.4e-33;  
 Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPFNIFSTNNLLKDIINFEYNNINDSKILSQNKNTLVDTSGYNVSEEGDVQLNP 62  
 Db 849 IPFOLSKYVNDQRLSTFTYIKNIINTSLNRYSNHLIDLSRYASKINGSKVNDP 908  
 Qy 63 IFPFD---FKLGSGEDRGKVIYTONENIYVNSWYSEFISFWIRINKVSNLP---GYT 116  
 Db 909 IDKQIQIQLFNLESS-----KIEVLKNAIYVNSWYSEFISFWIRINKVSNLP 963  
 Qy 117 IDSVKNNSGISGIIISNLFVLKQNEDEQISINFSYDISNNAPGY-NKWFVVTNNM 175  
 Db 964 IINCENNSGKVSGLNGYGEIITLQDTQEIQRVVFYKQMINISDIINRWIFVTIINR 1023  
 Qy 176 MGNKIYINGKLIDTIVKVELTGINFSTKTFEINKIPDTGLITSDSDNINMWIRDFYI 235  
 Db 1024 LNSKIYINGKLIDQIPISNLGNHASNIMFKLDCRDI-----HRYIWKYFNLF 1075  
 Qy 236 AKELDGKDINILFNSLQYTNVVDYNGNDRYNKEYYMWINDYLNMYMANS----- 287  
 Db 1076 DRELKEKEIKLDYDQNSGLKDFWGDYLDYQDPKPYMLNLYDPNKYVDVNNYVNGIRYMY 1135  
 Qy 288 ---RQIVENR-RNNDFNRYGKIIIRKIRGNTNDFVRGGDILYFDMTINKNKLFW 342  
 Db 1136 LKPGRSWTTNINSLYGTGTFIKKVASGNKDNIVANNDRVINVVYKKEYL-- 1193  
 Qy 343 KNETMYADNHSTEDIYAGLEQTKDINDNIIFQIOPMNNYYVYASQIFKSNFNGENISG 402  
 Db 1194 -----ATNASQAGVEKILSALEIPDVGNLSQVVMKSKNDQGITKRC-KMNLQDNNND 1246  
 Qy 403 ICSIGTYFRLGGDWYRHNLYVPTVKQGNVSLSTST---THWGFVPVSE 450  
 Db 1247 IGFIGFHQF-----NNIAKLVASNWNRYNRIQERSRILGCSWEIFPVD 1289

## RESULT 14

AA93308

ID AAY93308 standard; protein; 1070 AA.

XX AAY93308;

AC AAY93308;

XX 04-SEP-2000 (first entry)

DT A manganese superoxide dismutase (Mn-SOD) construct.

DE A manganese superoxide dismutase (Mn-SOD) construct.  
 XX Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;  
 KW neuronal cell targeting component; NCTC; neuronal disease;  
 KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;  
 KW Huntington's disease; motor neuron disease;  
 KW botulinum neurotoxin serotype B.

XX Synthetic.

OS *Bacillus stearothermophilus*.OS *Clostridium botulinum*.

XX WO200028041-A1.

PN 18-MAY-2000.

PD 05-NOV-1999; 99WO-GB03699.

XX AAY93311

RESULT 15

AA93311

XX 05-NOV-1998; 98GB-0024282.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.

XX Shone CC, Sutton JM, Hallis B, Silman N;

XX MPI; 2000-376553/32.

XX Novel composition, comprising superoxide dismutase linked by a  
 PT cleavable linker to a neuronal cell targeting component useful for  
 PT delivering superoxide dismutase to neuronal cells to treat ischemia -  
 XX Disclosure; Page 45-47; 65pp; English.

XX The present sequence represents a construct of the invention, comprising  
 CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that  
 CC can be cleaved by thrombin, and a heavy chain derived from botulinum  
 CC neurotoxin serotype B. The specification describes a composition for  
 CC delivery of SOD to neuronal cells. The composition comprises SOD linked,  
 CC by a cleavable linker, to a neuronal cell targeting component (NCTC).  
 CC This component has a domain that binds to a neuronal cell and a  
 CC domain that translocates the SOD of the composition into the neuronal  
 CC cell. After translocation, the linker is cleaved to release the SOD.  
 CC The composition is useful for treating neuronal diseases caused or  
 CC augmented by oxidative stress, such as ischemic stroke, trauma,  
 CC Parkinson's disease, Huntington's disease and motor neuron diseases.

XX Sequence 1070 AA;

Query Match 24.3%; Score 581; DB 21; Length 1070;

Best Local Similarity 32.9%; Pred. No. 2.8e-33;

Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;

Qy 3 IPFNIFSTNNLLKDIINFEYNNINDSKILSQNKNTLVDTSGYNVSEEGDVQLNP 62  
 Db 615 MPFDLSITVNTDILIEFNKYNSEILNLIILRYKDNLDLSGYGAKVYDVGELND 674  
 Qy 63 IFPFDKLGSGEDRGKVIYTONENIYVNSWYSEFISFWIRINKW---VSNL--PGYT 116  
 Db 675 --KNQFKLTSA--NSKIRVTONIIFNSVFLDFSVFWIRIPKYNKGDIQNYIHNEYT 730  
 Qy 117 IDSVKNNSGISGIIISNLFVLTKQNEDEQISINFSYDISNNAPGY-NKWFVVTNNM 175  
 Db 731 IINCENNSGKVSIRGNRIITLIDINGKTSVFEYNIREDISEYINRWFFVTITNN- 789  
 Qy 176 MGNKIYINGKLIDTIVKVELTGINFSTKTFEINKIPDTGLITSDSDNIN-MWIRDFYI 234  
 Db 790 LNNAKIYINGKLESNTDIKDIREVIANGEIIFKLD-----GDIDRTQIFWYFSTI 840  
 Qy 235 FAKELDGKDINILFNSLQYTNVVDYNGNDRYNKEYYMWINDYLNRYMYANS-----RQI 290  
 Db 841 FNTLSQSNIERYKIQSSEYKDFWGNPLMYNKEYYMFNAGNKNYSYIKLKSDSPVGEI 900  
 Qy 291 VFNTRNNN-----DFNEGYKIIIRKIRGNT---NDTRVRGGDILYFD-MTINKN--- 336  
 Db 901 LTRSKYNQNSKIYINRDLYIGKEFIIR-RKSNQSOSINDIVRVEDIYILFFNLDGWRV 959  
 Qy 337 -AYNLFEMKNETMYADNHSTEDIYAGLEQTKDINDNIIFQIOPMNNYYVYASQ-IFKSN 394  
 Db 960 YTYKFKKEE-----EKLFLAPISDSDEFYN---TIQKEYDEQPTSCQLLFFK- 1006  
 Qy 395 FNGENISGICSIGYRFRLLGG-----DWRHNYLVPTVKQGNVSLSTST 441  
 Db 1007 -DEESTDEIGLIGHRFYESGIVFEYKDYFCISKWY-----LKEVRKRPYKLNK---GC 1057  
 Qy 442 HWGFVPVSE 450  
 Db 1058 NWQFIPKDE 1066



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:43:01 ; Search time 26 Seconds  
(without alignments)  
509.243 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 WTTFNIFSTNNSLLKDI.....NYASLLESTSHWGTFVPS 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	586.5	24.5	1296	1	US-08-480-604A-28
2	586.5	24.5	1296	2	US-08-405-496A-28
3	586.5	24.5	1296	4	US-08-915-136-28
4	557.5	23.3	438	1	US-08-604A-23
5	557.5	23.3	438	2	US-08-405-496A-23
6	557.5	23.3	438	4	US-08-915-136-23
7	557.5	23.3	462	1	US-08-480-604A-26
8	557.5	23.3	462	2	US-08-405-496A-26
9	557.5	23.3	462	4	US-08-915-136-26
10	539.5	22.6	1169	4	US-08-235-829-20
11	425	17.8	853	4	US-08-913-880C-17
12	425	17.8	858	4	US-08-913-880C-16
13	425	17.8	860	4	US-08-913-880C-15
14	425	17.8	862	4	US-08-913-880C-14
15	425	17.8	865	4	US-08-913-880C-13
16	425	17.8	866	4	US-08-913-880C-12
17	425	17.8	874	4	US-08-913-880C-11
18	425	17.8	875	4	US-08-913-880C-10
19	425	17.8	1315	1	US-08-913-880C-1
20	423.5	17.7	618	1	US-08-668-381A-5
21	420	17.6	452	1	US-07-618-312A-2
22	420	17.6	452	1	US-07-618-312A-4
23	420	17.6	452	1	US-08-280-228-2
24	420	17.6	452	1	US-08-280-228-4
25	397	16.6	452	1	US-08-110-786A-8
26	148.5	6.2	977	4	US-09-206-942-53
27	148.5	6.2	983	4	US-09-206-942-51

28 143.5 6.0 599 2 US-08-910-551B-2 Sequence 2, Appli  
29 142.5 6.0 476 4 US-09-316-083-3 Sequence 3, Appli  
30 142.5 6.0 3135 1 US-08-323-170B-2 Sequence 2, Appli  
31 142.5 6.0 3135 4 US-08-954-441-2 Sequence 2, Appli  
32 138.5 5.8 2366 1 US-08-480-604A-10 Sequence 10, Appli  
33 138.5 5.8 2366 2 US-08-405-496A-10 Sequence 10, Appli  
34 138.5 5.8 2366 4 US-08-915-136-10 Sequence 10, Appli  
35 138.5 5.8 2366 4 US-08-957-310-10 Sequence 10, Appli  
36 128 5.4 1430 3 US-09-008-173-2 Sequence 2, Appli  
37 128 5.4 1430 4 US-09-210-361-6 Sequence 6, Appli  
38 128 5.4 1430 4 US-09-740-274-6 Sequence 6, Appli  
39 127.5 5.3 1529 2 US-08-728-470-10 Sequence 10, Appli  
40 127.5 5.3 1529 4 US-08-719-641-10 Sequence 10, Appli  
41 127.5 5.3 1545 4 US-08-236-791-4 Sequence 4, Appli  
42 127.5 5.3 1545 5 PCT-US95-10661A-4 Sequence 4, Appli  
43 127 5.3 1702 4 US-08-236-791-5 Sequence 5, Appli  
44 127 5.3 1702 5 PCT-US95-10661A-5 Sequence 5, Appli  
45 126 5.3 2710 1 US-08-480-604A-6 Sequence 6, Appli

## ALIGNMENTS

RESULT 1  
US-08-480-604A-28  
; Sequence 28, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/480,604A  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01763  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-480-604A-28

Query Match 24.5%; Score 586.5; DB 1; Length 1296;  
Best Local Similarity 31.6%; Pred. No. 2.6e-39;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPFNIFSYNNLSKLDIINFYNNINDSKILSONKKNLVDTSYNAVSEGDVQLNP 62  
Db 849 IPFOLSKYVDNQRLSFTFEIKNIINTSILNRYESNHLIDLSRYASKINSGKVNFD 908  
Qy 63 IFPPD---FKLGSSGDRGKVIYVTONENIVYNSMYESFSISFWIRINKVSNLP---GYT 116  
Db 909 IDKQIQIOLFNESS-----KIEVLKNAIVYNSMYENFSTFWIRIPKYPFNSISLNNEYT 963  
Qy 117 IDSVKNNSGWSIGIISNFTLVFTLKQNEDESEQSINFSYDISNNAPGY-NKWFVFTVNNM 175  
Db 964 IINCMMNSGKVSILYGEIITWLTQDTQEIQRVWFKYSQMINISDIYINRWIFVTITNR 1023  
Qy 176 MGNMKIYINGKLIDTIKVKELTGINSKTIITFEINKIPDTGLTSDNINMWIRDFYIF 235  
Db 1024 LNNKIIYINGKLIDQKPIISNLGNHASNNIMFKDGRDT-----HRYIWKYFNLF 1075  
Qy 236 AKELDGKDINILFNSLOYTNVVDYNGNDLRYNKYMYNIDYLNRYMYANS----- 287  
Db 1076 DKELNEKEIKDLYDNQNSGILKDFWGDYLOYDKPYMNLNLYDPNKYVDVNVGIRGYM 1135  
Qy 288 ----RQIVFNTR-RNNDFNEGKIIIKIRGTNTNTRVGGDILYFDMTINNKAYNLFM 342  
Db 1136 LKGRGVSMTNIIYNSLSYRGTRFKIIRKVASGNKNIVNNDRVINVVVNKEYRL-- 1193  
Qy 343 KNETMYADNHSTEDIYAIGLREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISG 402  
Db 1194 -----ATNASQAGVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGND 1246  
Qy 403 ICSTGTFRFLGGDWYRHNLYVTVKOGNTASLLEST-----THWGFVPVSE 450  
Db 1247 IGFIFGHQF-----NNIAKLVASNWNRYNRQIERSRSLTGLCSWEFIPVDD 1289

RESULT 2  
US-08-405-496A-28  
Sequence 28, Application US/08405496A  
Patent No. 5919865  
GENERAL INFORMATION:  
APPLICANT: WILLIAMS, JAMES A.  
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
TITLE OF INVENTION: NEUROTOXIN  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,496A  
FILING DATE: 16-MAR-1995

CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-405-496A-28

Query Match 24.5%; Score 586.5; DB 2; Length 1296;  
Best Local Similarity 31.6%; Pred. No. 2.6e-39;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

Qy 3 IPFNIFSYNNLSKLDIINFYNNINDSKILSONKKNLVDTSYNAVSEGDVQLNP 62  
Db 849 IPFOLSKYVDNQRLSFTFEIKNIINTSILNRYESNHLIDLSRYASKINSGKVNFD 908  
Qy 63 IFPPD---FKLGSSGDRGKVIYVTONENIVYNSMYESFSISFWIRINKVSNLP---GYT 116  
Db 909 IDKQIQIOLFNESS-----KIEVLKNAIVYNSMYENFSTFWIRIPKYPFNSISLNNEYT 963  
Qy 117 IDSVKNNSGWSIGIISNFTLVFTLKQNEDESEQSINFSYDISNNAPGY-NKWFVFTVNNM 175  
Db 964 IINCMMNSGKVSILYGEIITWLTQDTQEIQRVWFKYSQMINISDIYINRWIFVTITNR 1023  
Qy 176 MGNMKIYINGKLIDTIKVKELTGINSKTIITFEINKIPDTGLTSDNINMWIRDFYIF 235  
Db 1024 LNNKIIYINGKLIDQKPIISNLGNHASNNIMFKDGRDT-----HRYIWKYFNLF 1075  
Qy 236 AKELDGKDINILFNSLOYTNVVDYNGNDLRYNKYMYNIDYLNRYMYANS----- 287  
Db 1076 DKELNEKEIKDLYDNQNSGILKDFWGDYLOYDKPYMNLNLYDPNKYVDVNVGIRGYM 1135  
Qy 288 ----RQIVFNTR-RNNDFNEGKIIIKIRGTNTNTRVGGDILYFDMTINNKAYNLFM 342  
Db 1136 LKGRGVSMTNIIYNSLSYRGTRFKIIRKVASGNKNIVNNDRVINVVVNKEYRL-- 1193  
Qy 343 KNETMYADNHSTEDIYAIGLREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISG 402  
Db 1194 -----ATNASQAGVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGND 1246  
Qy 403 ICSTGTFRFLGGDWYRHNLYVTVKOGNTASLLEST-----THWGFVPVSE 450  
Db 1247 IGFIFGHQF-----NNIAKLVASNWNRYNRQIERSRSLTGLCSWEFIPVDD 1289

RESULT 3  
US-08-915-136-28  
Sequence 28, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.

APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/915,136  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1296 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-915-136-28

Query Match 24.5%; Score 586.5; DB 4; Length 1296;  
Best Local Similarity 31.6%; Pred. No. 2.6e-39;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

QY 3 IPFNFTNNSLLKDIINEFNNINSKILSLQNRKNTLVDSGYNAEVSSEGDVQLNP 62  
DB 849 IPQLSKYDQRLSTTTEYKNIINTSILNRYENHLIDLSRYASKINIGSKVNDP 908  
QY 63 IPFFD---FKLGSSGDRGVIVTQNEINIVNMYESFSFWRINKWVSNLP---GYT 116  
DB 909 IDKNOIQFNLSS-----KEVILKNAIVNSMYEFSFWRIRPKYNSISLNEYT 963  
QY 117 IIDSYKNSNGSIGIISNLFYFTLKQNEDSQSINFYDISNNAPGY-NKWFYFTVNNM 175  
DB 964 IINCENNSGKWSLVNLYGELIITLQDTEIKQVRVVKYSQMINISDYINRWIFVTINNR 1023  
QY 176 MGNMKIYINGKLIDIKYKELTGINFSKTITFEINKIPDPTGLTSDSDNINMWLRDPIYF 235  
DB 1024 LNNSKIYINGRLIDQKPISNLGNHASNNIMFKLDGRDT-----HRYIWKYFNLF 1075

QY 236 AKELDGKDINILFNSLOYTNNVKDYKGNDRYNKNEYMMNIDLYNRYMYANS----- 287  
DB 1076 DKLENEKEIKDLYDNQNSGILKDFWDYLOYPKPYMLNLYDPNKYVDNKNVIRGYMY 1135  
QY 288 ---RQIVFNR-RNNNDFEGYKIIKIRKRGNTNDRVRGGDILYFDMTINNKAYNLFM 342  
DB 1136 LKPRGSGVMTNIVLNSSLYRGKFKFIKKYASGNKNIVRNRDRVINNVYKKEVRL-- 1193  
QY 343 KNETWYADNHSTEDIYAIGLRBQTKDINDNIIQIOPMNTYYASQIFKSPNGENISG 402  
DB 1194 -----ATNASQACVEKILSALEIPDVGNLSQVVVVKSKNDQGITNKC-KMNLQDNNGND 1246  
QY 403 ICSIGTYRFRLLGGDWYRHNYLVPVTKQGNVYASLLEST-----THWGFVPVSE 450  
DB 1247 IGFIGFHFQF-----NNIAKLIVASNNWNRQIERSRRLGCSWFEFIPVDD 1289

RESULT 4  
US-08-480-604A-23  
; Sequence 23, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 23:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 438 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-480-604A-23

Query Match 23.3%; Score 557.5; DB 1; Length 438;  
Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

QY 16 LKDIINEYFNINDSKILSLQNKRLTVDTSYNAEVSSEGVDQLNPFPFD---FKLGS 72  
DB 4 LLSTFEYIKNIINTSLNRLYESHLIDLSRYASKINGSKVNFDPDKNQIQLFNLES 63  
QY 73 SGEDRGKVIIVTQENIVYNSMYESFSISFWIRKINKVSNLP---GYTIIDSVKYNSGHSI 129  
DB 64 S-----KIEVILKNAIVNSMYENFSTFWIRPKYFNSISLNNEYTIINCNNNSGKW 118  
QY 130 GIISNLFVFLTKONESEQISNFSDISNAPGY-NKWFVTVTNMGMKVIYINGKLI 188  
DB 119 SLNYGEIITWLTQTEIKQVYKYSQMINISDYINRWIFVITNNLNNKIYINGRLI 178  
QY 189 DTIKVKELTGINFSTKITFEINKIPDTGLTSDSDNNINWIRDFYIFAKELDGKINILF 248  
DB 179 DQKPISNLGNIHASNNIMFKLDCRDT-----HRYIWKYFNLFDKELNEKEIKDLY 230  
QY 249 NSLQYTNVVDYNGDRLRYNKYEYVYVNIIDYLNRYMYANS-----RQIVFENTR- 295  
DB 231 DNQNSGILKDFWGDYLDQYDKPYMLNLDPNKYVDVNVNGVGYLKGPRGSVWTTNI 290  
QY 296 RNNDFNEGKIIKRIKRGNTNDRVRGGDILYFDMTINKKAYNLFMKNETWYADNHSTE 355  
DB 291 YLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVKNKEYRL-----ATNASQA 342  
QY 356 DIVAIGLREQTKDINDNIIFQIOPMNTYVYASQIFKSNFNGENISGICSGTYRFLGG 415  
DB 343 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNNDIGFIGHQF----- 397  
QY 416 DWYRHNYLVPTVKOGNYASLLESTST---THWGFVPVSE 450  
DB 398 -----NNIAKIVASNMWNRQIERSRSLTGCSEWEIFPVDD 431

RESULT 5  
US-08-405-496A-23  
; Sequence 23, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; MEDIUM OF INVENTION: NEUROTOXIN  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR-APPLICATION DATA:

; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01308  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 438 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-405-496A-23

Query Match 23.3%; Score 557.5; DB 2; Length 438;  
Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

QY 16 LKDIINEYFNINDSKILSLQNKRLTVDTSYNAEVSSEGVDQLNPFPFD---FKLGS 72  
DB 4 LLSTFEYIKNIINTSLNRLYESHLIDLSRYASKINGSKVNFDPDKNQIQLFNLES 63  
QY 73 SGEDRGKVIIVTQENIVYNSMYESFSISFWIRKINKVSNLP---GYTIIDSVKYNSGHSI 129  
DB 64 S-----KIEVILKNAIVNSMYENFSTFWIRPKYFNSISLNNEYTIINCNNNSGKW 118  
QY 130 GIISNLFVFLTKONESEQISNFSDISNAPGY-NKWFVTVTNMGMKVIYINGKLI 188  
DB 119 SLNYGEIITWLTQTEIKQVYKYSQMINISDYINRWIFVITNNLNNKIYINGRLI 178  
QY 189 DTIKVKELTGINFSTKITFEINKIPDTGLTSDSDNNINWIRDFYIFAKELDGKINILF 248  
DB 179 DQKPISNLGNIHASNNIMFKLDCRDT-----HRYIWKYFNLFDKELNEKEIKDLY 230  
QY 249 NSLQYTNVVDYNGDRLRYNKYEYVYVNIIDYLNRYMYANS-----RQIVFENTR- 295  
DB 231 DNQNSGILKDFWGDYLDQYDKPYMLNLDPNKYVDVNVNGVGYLKGPRGSVWTTNI 290  
QY 296 RNNDFNEGKIIKRIKRGNTNDRVRGGDILYFDMTINKKAYNLFMKNETWYADNHSTE 355  
DB 291 YLNSSLYRGTKFIKKYASGNKDNIVRNDRVYINVVKNKEYRL-----ATNASQA 342  
QY 356 DIVAIGLREQTKDINDNIIFQIOPMNTYVYASQIFKSNFNGENISGICSGTYRFLGG 415  
DB 343 GVEKILSALEIPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNNDIGFIGHQF----- 397  
QY 416 DWYRHNYLVPTVKOGNYASLLESTST---THWGFVPVSE 450  
DB 398 -----NNIAKIVASNMWNRQIERSRSLTGCSEWEIFPVDD 431

RESULT 6  
US-08-915-136-23  
; Sequence 23, Application US/08915136  
; Patent No. 6290960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHIE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: MEDLEN & CARROLL, LLP  
;; STREET: 220 MONTGOMERY STREET, SUITE 2200  
;; CITY: SAN FRANCISCO  
;; STATE: CALIFORNIA  
;; COUNTRY: UNITED STATES OF AMERICA  
;; ZIP: 94104  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/915,136  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/480,604  
;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/405,496  
;; FILING DATE: 16-MAR-1995  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/329,154  
;; FILING DATE: 25-OCT-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/161,907  
;; FILING DATE: 02-DEC-1993  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/985,321  
;; FILING DATE: 04-DEC-1992  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 07/429,791  
;; FILING DATE: 31-OCT-1989  
;; NAME: INGOLIA, DIANE E.  
;; ATTORNEY/AGENT INFORMATION:  
;; REGISTRATION NUMBER: 40,027  
;; REFERENCE/DOCKET NUMBER: OPD-01763  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 705-8410  
;; TELEFAX: (415) 397-8338  
;; INFORMATION FOR SEQ ID NO: 23:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 438 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-915-136-23

Query Match 23.3%; Score 557.5; DB 4; Length 438;  
Best Local Similarity 31.4%; Pred. No. 1.3e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;  
QY 16 LKDIINEYFNINDSKILSLQNKNTLVDTSYNAEVSSEGDVQLNPFFPD---FKLGS 72  
Db 4 LLSTFEYIKNIINTSLNLRYESNHLDSLRYASKINIGSKVNTFPDIDKQIQLFNLES 63  
QY 73 SGDEKGVITQENIVYMSVESPISFWIRKWSNLP---GYTIIDSKVKNNSGWSI 129  
Db 64 S-----KIEVLKNAIVYMSVENSFSEWIRPKYFNSISLNNSEYITINCMMNNSGKV 118  
QY 130 GIISNPLVFTLKNEDSEGISNFSYDISNAPGY-NKWFVFTVITNNMKNKIYINGKLI 188  
Db 119 SLNYGEIITLQDTQEKORVVKFSQMINISDYINRWITFTITNRLNNSKIYINGLI 178  
QY 189 DTIKVELTGINSKTTTEINKIPDTGLITSDSNININWIRDFYFAKELGDKDINILF 248  
Db 179 DQKPIISNLGNIHASNNIMPKLDCRDT-----HRYTIWKYFNLFKELNEKEIKOLY 230  
QY 249 NSLQYTNVYKWDGNDLRYNKEYYMYNIDLYNRYVANS-----RQIVFNIR- 295  
Db 231 DNASGILKDFWGLDYLQDKPIYMLNLYDPNRYVDNNVNGIRGMYLKGPRGVSMTINI 290

QY 296 RNNDFNEGYKIIIRKRGNTNDRVRCGDILYFDMTINNKAYNLFMKNETMYADNHSTE 355  
Db 291 YLNSSLYRGTKFIIRKYASGNKNVIRNRDRIYINVVYKNEYRL-----ATNASQA 342  
QY 356 DIVAIGLREOTKDINDNIIFQIPMNNTYVYASQIFKNGFNNGENISGICSGTGYRFLGG 415  
Db 343 GVEKILSALEIFDVGNLQSVVYKSKNDQGITNKC-KNNLQDNNGNDIGFIGFHQF---- 397  
QY 416 DMRYRHNLYVTYKQGNYSALLESTS-----THWGFVPVSE 450  
Db 398 -----NNIAKLVSANWYNRQIERSRSLTLCGSWEFIPVDD 431  
RESULT 7  
US-08-480-604A-26  
; Sequence 26, Application US/08480604A  
; Patent No. 5736139  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,604A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/422,711  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 26:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-480-604A-26

Query Match 23.3%; Score 557.5; DB 1; Length 462;  
Best Local Similarity 31.4%; Pred. No. 1.4e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

QY 16 LKDIINEYNNINDSKILSLQNRKNTLVDTSGYNRAVSEEGDVLNPIFPD---FKLGS 72  
DB 28 LLSTFTEYIKNIINTSILNRYESNHLIDLRYASKINIGSKVNFDPDKNQIOFLNLES 87  
QY 73 SGEDRGKVIQTQENIVNYSMEYSFISFWIRINKWVSNLP---GYTIIDSVKNNSGWSI 129  
DB 88 S-----KIEVILKNAIVNYSMEYENSTFWIRIPKPFNSISLNNEYTIINCMMENSGWKV 142  
QY 130 GIISNPLVTLKQNEDESGINSFSDISNNAPGY-NKWFVTVTNMMGNKIYINGKLI 188  
DB 143 SLNYGEIITLQDTQIKORVVKYQSMINISDYINRWIFVTITNNRLNNSKIYINGRLI 202  
QY 189 DTIKVRELGTINFSKTIPTFEINKIPDTGLITSDSNINKNWIRDFYIFAKELDGKDINILF 248  
DB 203 DQKPISNLGNIHASNNIMEKLDGCRDT-----HRYIWIKYFLNFKELNEXEIKDLY 254  
QY 249 NSLOQYTNVYKDYNGNDLRYNKYKYYMNDILNRYMYANS-----RQIVFNTR- 295  
DB 255 DNGNSGILKDFWGDYLDYKPYMLNLYDPNKYVDVNVGIRGYMYLKGPRGSMYMTINI 314  
QY 296 RNNNDFNEGYKIIKIRGNTNDTRVGGDILYFDMTNNKAYNLFEMKNETMYADNHSTE 355  
DB 315 YLNSSLYRGTKFIKKYASGKNDNVNRNDRYINVVYKNKEYRL-----ATNASQA 366  
QY 356 DIVAIGLREQTKDINDNIIFQIPMNNITYYASQIFKSNFNGENISGICSTGTVRFRLLGG 415  
DB 367 GVEKILSALEIPDVGNLSQVVMYKSKNDQGITNKC-KMNLQDNNGNDIGFICGHQF---- 421  
QY 416 DWYRHNYLVPTVKQGYASLLEST-----THWGFYVSE 450  
DB 422 -----NNIAKLVASNWNRYNQIERSSRTLCGSWEFIPVDD 455

## RESULT 8

US-08-405-496A-26  
Sequence 26, Application US/08405496A  
Patent No. 5919665  
GENERAL INFORMATION:  
APPLICANT: WILLIAMS, JAMES A.  
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
TITLE OF INVENTION: NEUROTOXIN  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405,496A  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 26:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 462 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-403-496A-26

Query Match 23.3%; Score 557.5; DB 2; Length 462;  
Best Local Similarity 31.4%; Pred. No. 1.4e-37;  
Matches 144; Conservative 80; Mismatches 180; Indels 55; Gaps 11;

QY 16 LKDIINEYNNINDSKILSLQNRKNTLVDTSGYNRAVSEEGDVLNPIFPD---FKLGS 72  
DB 28 LLSTFTEYIKNIINTSILNRYESNHLIDLRYASKINIGSKVNFDPDKNQIOFLNLES 87  
QY 73 SGEDRGKVIQTQENIVNYSMEYSFISFWIRINKWVSNLP---GYTIIDSVKNNSGWSI 129  
DB 88 S-----KIEVILKNAIVNYSMEYENSTFWIRIPKPFNSISLNNEYTIINCMMENSGWKV 142  
QY 130 GIISNPLVTLKQNEDESGINSFSDISNNAPGY-NKWFVTVTNMMGNKIYINGKLI 188  
DB 143 SLNYGEIITLQDTQIKORVVKYQSMINISDYINRWIFVTITNNRLNNSKIYINGRLI 202  
QY 189 DTIKVRELGTINFSKTIPTFEINKIPDTGLITSDSNINKNWIRDFYIFAKELDGKDINILF 248  
DB 203 DQKPISNLGNIHASNNIMEKLDGCRDT-----HRYIWIKYFLNFKELNEXEIKDLY 254  
QY 249 NSLOQYTNVYKDYNGNDLRYNKYKYYMNDILNRYMYANS-----RQIVFNTR- 295  
DB 255 DNGNSGILKDFWGDYLDYKPYMLNLYDPNKYVDVNVGIRGYMYLKGPRGSMYMTINI 314  
QY 296 RNNNDFNEGYKIIKIRGNTNDTRVGGDILYFDMTNNKAYNLFEMKNETMYADNHSTE 355  
DB 315 YLNSSLYRGTKFIKKYASGKNDNVNRNDRYINVVYKNKEYRL-----ATNASQA 366  
QY 356 DIVAIGLREQTKDINDNIIFQIPMNNITYYASQIFKSNFNGENISGICSTGTVRFRLLGG 415  
DB 367 GVEKILSALEIPDVGNLSQVVMYKSKNDQGITNKC-KMNLQDNNGNDIGFICGHQF---- 421  
QY 416 DWYRHNYLVPTVKQGYASLLEST-----THWGFYVSE 450  
DB 422 -----NNIAKLVASNWNRYNQIERSSRTLCGSWEFIPVDD 455

## RESULT 9

US-08-915-136-26  
Sequence 26, Application US/08915136  
Patent No. 6290960  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA







[illegible]

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

1

Db 574 NAYLANKWFTITNDRLSSANLYINGVLMSGAEITGLGAIRDNNITLKLDL----- 626  
QY 219 TSDSDNINMWRDFYIFAKELDKDINILFNSLQYTNVVKDYWGNDLRYNKEYYKYN-- 276  
Db 627 -CANNNGYVSIDKFRIFCKALNPKETKLYTSYLSITFLRDFWGNPLRYDTEYLLIPVAS 685  
QY 277 -----DYL--NRYMYANSRQIVENTRRNNDFNEGYKIIKKIRGNTN-DREV 321  
Db 686 SSKDVQLKNITDYMILTNAPSYNGKLNYYRRLN-----GLKFIKRYTPNNEIDSFV 740  
QY 322 RGGDILYFDMTINNKA-----NLFMKNETMYADNHSTEDIYAIGLREQTKDINDNI 373  
Db 741 KSGDFIKLYSVSYNNNEHIVGPKDGNFNNLDRILRVGYNAPGIPLYKKMEAVK----- 794  
QY 374 IFQIQPMNNTYYIASQIFKSNFNGENISGICSIGTYRFRLGSD-----WYRHNVL 423  
Db 795 -----LRDLKTSVQL--KLYDDKNAS-LGLVGTNGQIGNDPNDRDILIASNWY-FNEL 844  
QY 424 VPTVKOGNYASLLESTSTHMGFVPVSE 450  
Db 845 KDKI-----LGCDWYFVPTDE 860

Search completed: November 7, 2002, 14:47:30  
Job time : 30 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:45:51 ; Search time 19 seconds  
(without alignments)  
341.491 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIPFNFSYNNLLKDI.....NYASLLESTHGWGPVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 92612 seqs, 14418503 residues

Total number of hits satisfying chosen parameters: 92612

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	586.5	24.5	1295	10 US-09-726-949A-1	Sequence 1, Appli
2	546	22.8	431	8 US-08-981-087A-1	Sequence 1, Appli
3	544.5	22.8	425	10 US-09-288-326-9	Sequence 9, Appli
4	250	10.5	144	8 US-08-981-087A-2	Sequence 2, Appli
5	239	10.0	144	8 US-08-981-087A-3	Sequence 3, Appli
6	145.5	6.1	1974	9 US-09-895-913A-12	Sequence 12, Appli
7	142.5	6.0	476	10 US-09-774-414-3	Sequence 3, Appli
8	133	5.6	932	10 US-09-815-242-5578	Sequence 5578, Ap
9	133	5.6	932	10 US-09-815-242-12438	Sequence 12438, A
10	129	5.4	841	10 US-09-815-242-5779	Sequence 5779, Ap
11	129	5.4	841	10 US-09-815-242-12751	Sequence 12751, A
12	128	5.4	1430	10 US-09-740-274-6	Sequence 6, Appli
13	121.5	5.1	990	12 US-10-047-678A-7	Sequence 7, Appli
14	121	5.1	483	8 US-08-834-666A-20	Sequence 20, Appli
15	118	4.9	993	10 US-09-815-242-5809	Sequence 5809, Ap
16	118	4.9	1002	10 US-09-815-242-12899	Sequence 12899, A
17	118	4.9	1002	10 US-09-815-242-13158	Sequence 13158, A
18	116.5	4.9	515	10 US-09-925-300-1282	Sequence 1282, Ap
19	116.5	4.9	996	10 US-09-815-242-5251	Sequence 5251, Ap

20	116.5	4.9	1009	10 US-09-815-242-12141	Sequence 12141, A
21	111.5	4.7	621	10 US-09-856-247A-2	Sequence 2, Appli
22	111.5	4.7	789	10 US-09-995-587A-1	Sequence 1, Appli
23	110.5	4.6	586	10 US-09-861-451A-32	Sequence 32, Appli
24	110	4.6	815	10 US-09-815-242-5106	Sequence 5106, Ap
25	110	4.6	824	10 US-09-866-582-34	Sequence 34, Appli
26	110	4.6	1116	10 US-09-790-318-2	Sequence 2, Appli
27	108	4.5	1437	10 US-09-801-368-354	Sequence 354, App
28	107	4.5	626	10 US-09-765-272-220	Sequence 220, App
29	107	4.5	869	10 US-09-815-242-10623	Sequence 10623, A
30	106.5	4.5	1781	9 US-09-995-749A-2	Sequence 2, Appli
31	106	4.4	691	8 US-08-834-666A-12	Sequence 12, Appli
32	106	4.4	691	8 US-08-834-666A-22	Sequence 22, Appli
33	104.5	4.4	555	8 US-08-808-031A-31	Sequence 31, Appli
34	104.5	4.4	586	8 US-08-808-031A-28	Sequence 28, Appli
35	103.5	4.3	584	9 US-09-995-749A-12	Sequence 12, Appli
36	103	4.3	1093	10 US-09-801-368-352	Sequence 352, App
37	102.5	4.3	1579	10 US-09-801-368-368	Sequence 368, App
38	101.5	4.2	476	10 US-09-733-524-15	Sequence 15, Appli
39	101.5	4.2	792	10 US-09-995-587A-11	Sequence 11, Appli
40	101	4.2	283	10 US-09-846-808-7	Sequence 7, Appli
41	101	4.2	283	10 US-10-059-964-26	Sequence 26, Appli
42	101	4.2	660	10 US-09-881-752A-22	Sequence 22, Appli
43	101	4.2	1143	10 US-09-924-154-14	Sequence 14, Appli
44	100.5	4.2	886	9 US-10-002-309B-2	Sequence 2, Appli
45	100.5	4.2	1031	10 US-09-815-242-10932	Sequence 10932, A

## ALIGNMENTS

RESULT 1  
US-09-726-949A-1  
; Sequence 1, Application US/09726949A  
; Patent No. US20020137886A1  
; GENERAL INFORMATION:  
; APPLICANT: Allergan, Inc.  
; APPLICANT: Lin, Wei-Jen  
; APPLICANT: Aoki, Kei Roger  
; APPLICANT: Steward, Lance E.  
; TITLE OF INVENTION: NEUROTOXINS WITH ENHANCED TARGET  
; TITLE OF INVENTION: SPECIFICITY  
; FILE REFERENCE: 36121-20002.00  
; CURRENT APPLICATION NUMBER: US/09/726,949A  
; CURRENT FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1295  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-09-726-949A-1

Query Match 24.5%; Score 586.5; DB 10; Length 1295;  
Best Local Similarity 31.6%; Pred. No. 4.3e-37;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;  
Qy 3 IPENFSTNNLLKDIINEYFNNDKSKILQNRKNTLVDTSGYNAEYSEGDYQLNP 62  
Db 848 IPFQLSKVDNQRLLSTFTTEKINILNRYESNHLIDLSRYASKINIGSKYNEFP 907  
Qy 63 IFPFD---FKIGSSGDEGRKVIYQENINIVNSMYEFSFIRINKWVSNLP---GYT 116  
Db 908 IDKNGIQLFNLESS-----KIEVLKNAIVNSMYENFSTFIRIPKYPNSLNNEXT 962  
Qy 117 IIDSVKNSGSGSIGIISNFIIVTLKQNEDESEQISNFSDISNNAPGY-NKWFVTVTNM 175  
Db 963 INCMENNSGKVSGLNTGEIILWLTQDQEKQVVFYKYSOMINISDIYNRWIVFTNNR 1022  
Qy 176 MGNKIIYNGKLIDITIKVKELTGINFKTIIFFEINKIPDTGLITSDSNINMWIRDFYIF 235  
Db 1023 LNNKIIYNGRLIDQKPIISNLGNTHASNIMFKLDGCRDT-----HRYIWKYFNLF 1074



Db 228 DFWGDLQYDKPYMLNLDPNKYVDVNNVGRGYMLKPRGSVMTNIIYLNSSLYRGT 287  
QY 306 KILIKIRGNTDTRVGGDILYEDMTINNKAYNLFKNETYADNHSTEDIYAIGLREQ 365  
Db 288 KFIKKYASGNKONIVNRDRVINVVVRNKEYRL-----ATNASAGVEKILSALE 339  
QY 366 TKDINDNIIQIOPMNNYYASQIFKSNFNGENISGICISGTYFRGLGDWYRHNLYVP 425  
Db 340 IPDVGNLSQVVMKSKNDQGITNKC-KMNLQDNNGNDIGFIGHQF-----NNIAK 389  
QY 426 TVKQGNVASLLEST-----TWGFFVPVSE 450  
Db 390 LVASNWNQIERSRRLGCSWEFIPVDD 418

RESULT 4  
US-08-981-087A-2  
; Sequence 2, Application US/08981087A  
; Patent No. US20020081304A1  
; GENERAL INFORMATION:  
; APPLICANT: Elmore, Michael J.  
; APPLICANT: Mauchline, Margaret L.  
; APPLICANT: Minton, Nigel P.  
; APPLICANT: Pasechnik, Vladimir A.  
; APPLICANT: Titball, Richard W.  
; TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NO. US20020081304A1th Glebe Rd. 8th floor  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201-4741

; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/981,087A  
; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/01409  
; FILING DATE: 12-JUN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9511909.5  
; FILING DATE: 12-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crawford, Arthur R.  
; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 124-688  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 144 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-981-087A-2

Query Match 10.5%; Score 250; DB 8; Length 144;  
Best Local Similarity 37.7%; Pred. No. 1.3e-12;  
Matches 55; Conservative 30; Mismatches 55; Indels 6; Gaps 4;  
QY 9 SYTNLSLKDINEYFNNDKSLSKLONKRNKTLVDTSGYNAEVSSEGOVQLNPFPDF 68  
Db 1 SYTNKILILYFNKLYKIKDNLDRYENKFDISGYGNSINISGDIYIYSTNRNOF 60

QY 69 KLGSSGDRGKVIYTONENIVNMYESFISFWIRKWS--NLPG-YTIIDSVK-NN 124  
Db 61 GIYSS--KPEVNIAQNNDIYNGRYQNTSISFWAIPKYFNKVLNNNEYTIIDCIRNN 118  
QY 125 SGWSIGLISNLFVTLKQNEDESQSI 150  
Db 119 SGWKISLNYKIIITLODTAGNNQKL 144

RESULT 5  
US-08-981-087A-3  
; Sequence 3, Application US/08981087A  
; Patent No. US20020081304A1  
; GENERAL INFORMATION:  
; APPLICANT: Elmore, Michael J.  
; APPLICANT: Mauchline, Margaret L.  
; APPLICANT: Minton, Nigel P.  
; APPLICANT: Pasechnik, Vladimir A.  
; APPLICANT: Titball, Richard W.  
; TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHYE P.C.  
; STREET: 1100 NO. US20020081304A1th Glebe Rd. 8th floor  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201-4741  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/981,087A  
; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/01409  
; FILING DATE: 12-JUN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9511909.5  
; FILING DATE: 12-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crawford, Arthur R.  
; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 124-688  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 144 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-981-087A-3

Query Match 10.0%; Score 239; DB 8; Length 144;  
Best Local Similarity 37.0%; Pred. No. 9.1e-12;  
Matches 54; Conservative 25; Mismatches 57; Indels 10; Gaps 2;  
QY 152 PSYDISNNAPGY-NKFFVTVTNMGMKIIYNGKLIDTIKVELTGINFSTIIFEIN 210  
Db 2 FNYTQMISIDSYINKWIFVTITNNRNGSRIYINGNLDEKISINLGDHVSNNILFKIV 61  
QY 211 KIPDTGLITSDSDNNMWIRDFYIFAKELDKGKDNILFNSLQYTNVVKDYWGNDLRNKE 270  
Db 62 GCNDRYV-----GIRYFKVFOFELCKTELELYSDPEPDSILKDFWGNLYNKR 112  
QY 271 YMYNDYLNRYMYANSRQIVFNTR 296

Db 113 YVLLNLRDTSITQNSFNINQOR 138

RESULT 6

US-09-895-913A-12

; Sequence 12, Application US/09895913A

; Patent No. US20020160456A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles

; APPLICANT: Tomb, Jean Francois

; APPLICANT: Comen, Raymond P.

; TITLE OF INVENTION: Identification of Polynucleotides

; TITLE OF INVENTION: Encoding No. US20020160456A1el Helicobacter Polypeptides in the

; TITLE OF INVENTION: Genome

; FILE REFERENCE: 06132/043002

; CURRENT APPLICATION NUMBER: US/09/895,913A

; CURRENT FILING DATE: 2001-06-29

; PRIOR APPLICATION NUMBER: US 08/881,227

; PRIOR FILING DATE: 1997-06-24

; NUMBER OF SEQ ID NOS: 368

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12

; LENGTH: 1974

; TYPE: PRT

; ORGANISM: Helicobacter pylori

US-09-895-913A-12

Query Match 6.1%; Score 145.5; DB 9; Length 1974;

Best Local Similarity 21.2%; Pred. No. 0.003;

Matches 106; Conservative 61; Mismatches 143; Indels 191; Gaps 24;

QY 19 IINEYFNINDSKILSLQNRKNTLVDTSGYNAEVEGVDQLNPFPPDFKLGSSGDRG 78

Db 60 IHSLFHTDADSKDVSQVRKQ-----FDPIPKTEPCVG 94

QY 79 KVIIVT-QNEINYNMYSEFSISFWIRINKWNSLPGYTIIDSV-----KNSGWS 128

Db 95 VCYIAPYKNQDLIGSSAF-AWSLNF-----GATVGTLLLGSAGEKANNNGS 141

QY 129 IGIISNFVLF-----TLKQNEDESEQISNFSYDISNNAFPGYNKWFVTV 171

Db 142 IFWGNLLYLHGNFATNIFLTNNVGNPNAGGATINFNADETINADGLNNTNFQTV 201

QY 172 TNNMGNMKIY-----INGKLIDTK---VKELT--GINF-SKTIPE-----INKI 212

Db 202 ALGLQTSASQHSWANFNSKLSMEIKNSNFRDFTWGGFNFSGRITFENTFTSGWTNIN-- 259

QY 213 PDTGLITSDSONINWIRDFYIFAKELDGKDI-----NILFNSLQYT-----NYVKDY 260

Db 260 ---GATESGSSVNVAVNTDILFNSILGGIRYDLKANNIIFNNSQWIDVSKNVNQSS 316

QY 261 WNDLRYNKEYYMW-----NDYLN-----NYMY 284

Db 317 LAGNVTFNNSRLSVKPNAAINIGDSQOTQALENASSLSFYNNSVANPNGTAFNGVSYLN 376

QY 285 AN-SRQIVFTRNNN-----DENEGKIIKIRKRGNTNTRVRGGDILY 328

Db 377 LNPNAQVSNQVNFNANVTYGPFGKTFDFGNSARLI--NFKGNTN-----F 424

QY 329 FDMTINNKAYNL-----FMKNETMYADNHSTEDIYAGLEQTKDNDNIIFQIQP 379

Db 425 NQATNLRAKHINHFQGVSTFKQNSTMTNLAESSQAFNALKVGEET---NFN----- 474

QY 380 MNNTYYASQIFKSNFNGENI 400

Db 475 LNNSSL-----NFGNSV 488

RESULT 7

US-09-774-414-3

; Sequence 3, Application US/09774414

; Patent No. US20020102231A1

; GENERAL INFORMATION:

; APPLICANT: The Institute of Physical and Chemical Research

; TITLE OF INVENTION: Endonuclease

; FILE REFERENCE: PH-651

; CURRENT APPLICATION NUMBER: US/09/774,414

; CURRENT FILING DATE: 2001-01-31

; PRIOR APPLICATION NUMBER: 09/306,970

; PRIOR FILING DATE: 1999-05-07

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 476

; TYPE: PRT

; ORGANISM: Saccharomyces cerevisiae

US-09-774-414-3

Query Match 6.0%; Score 142.5; DB 10; Length 476;

Best Local Similarity 22.2%; Pred. No. 0.0089;

Matches 81; Conservative 58; Mismatches 105; Indels 121; Gaps 20;

QY 11 TNNSLKLD--IINEY-----FNNINDSKILSLQNRKNTLVDTSGYNAEVS 53

Db 148 SNFYLMKYLINKYKMYKYLDMNIPNMYNMYNNYKGLNIK---TVLDLN--NNEFY 201

QY 54 E-----EGDVQLNPFPDFKLGSSGDRGVIVTQENIVYNSMYESFSISFWIRNK 107

Db 202 DYLSGLIEGSGYIGP-----GGITIHANDVLTIFINKRKNSILVEK 246

QY 108 WSNLPGYTIIDSVKNN-----SGWSIGIISNEL---VFTLKONE-DSEQSINFSYDISNN 159

Db 247 W-----MDILKONPYFVNAFSINIKTLAKERIFITNINLYKLYSDYKIN--QINNH 294

QY 160 APGVNKKWFFVTYVN-----NMGNMKIY-----INGKLI-----DTIKVKE----- 195

Db 295 IPYYN---YLKINKNPKIKNMDIKNNYLAGTAADGSLSSWYNPKDTLLFKNNRPSY 351

QY 196 -LTGINFSKTIITEI-----NKIPDTGLITSDSDNINWIRDFYIFAKE 238

Db 352 VISQVTERKELIYLOBSFDSLSNVKVKGNRKLDKFLTRTTDELKMFYFYDFKFLPL 411

QY 239 LOGKDINLNSQYTNVNDYWGNDLRYNKEYYMYNIDYLN-----YMYANSRQIV 291

Db 412 HDNQFNYI--KERFNTFIKSYNWN---NRVGLVLYSEINNINIKINDYDYNNKYIN 465

QY 292 ENTRR 296

Db 466 HNARK 470

RESULT 8

US-09-815-242-5578

; Sequence 5578, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; TITLE OF INVENTION: Prokaryotes

; FILE REFERENCE: ELITRA.011a

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727



```
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5578
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5578

Query Match      5.6%; Score 133; DB 10; Length 932;
Best Local Similarity 17.6%; Pred. No. 0.011;
Matches 86; Conservative 85; Mismatches 159; Indels 160; Gaps 19;

QY 6 NIFSYNNSLLKDI-----INEFNNINDSKILSLQNRKNTLVDTSYNA 50
DB 250 NLYNAQGNIIAKGYDSETSTTFTTNYVDQYTNISGSFEQVAFAKRENATDKTAYPM 309
QY 51 EYSEEGD-VOLNPIPPDFKLGSSGDEGRKVIYTONENIVYNSMYESFSISFWIRNK-- 107
DB 310 EYVLGNDKYSKNVIVDY-----GNQKGQQLIS-STNYINN--EDLSRMTVYVYNQPK 358
QY 108 -----WVSNLPGYTTIDSVKNNSGWSIGIISNPL-VFTLKONEDSEQSINFSYDINN 159
DB 359 KTYTKETFTVTLTGKFNPDAKNKIYEVTNQNFVDSFTPTDKLTVDTKFKITYSND 418
QY 160 APGYNKWFVTVTNMGMNKIYI-----NGKLIDTIKVK-----ELTG 198
DB 419 ----NKTATVLLNGQSSDKQYIIQQVAYPDNSSTNGKIDYILETQNGKSSWSNSYN 474
QY 199 INFSTIITFEINKIPDTGLITSDS-----DNINMWIRDFYIFAKELDGKDN----- 245
DB 475 VNGSSTANGDQKKYNLGDIYVWEDTKDGKODANEKIGKGVYVILKDSNGKELDRITTDEN 534
QY 246 -----ILFNSLQYTNVVKDYWGNDLRYNKE 270
DB 535 GKYQFTGLNGTYSVEFTLAGYPTTANAGTDDAVIDSDGLTTGVKID--ADNMILDSG 592
QY 271 YVMNIDYLNRYMYANSR-----QIVFNTRNNN-----D 300
DB 593 FYKTPKYSILGDYVWYDSNKGDKQDSTEKGIDKVKVTLNNEKGEVIGTKTDENGKYRFDN 652
QY 301 FNEG-YKIIIRKIRGNT-----NDRVRGGDI-----LYFDMTINNKAYNLFMKNET 346
DB 653 LDSGKYKVIKFEKPAGLTGTGTNTTEDDKADGGEVDVITDHDHDFLLDNGYY-----EET 708
QY 347 MYADNHSTED 356
DB 709 SDSDSDSDD 718

RESULT 9
US-09-815-242-12438
; Sequence 12438, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
```

```
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12438
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12438

Query Match      5.6%; Score 133; DB 10; Length 932;
Best Local Similarity 17.6%; Pred. No. 0.011;
Matches 86; Conservative 85; Mismatches 159; Indels 160; Gaps 19;

QY 6 NIFSYNNSLLKDI-----INEFNNINDSKILSLQNRKNTLVDTSYNA 50
DB 250 NLYNAQGNIIAKGYDSETSTTFTTNYVDQYTNISGSFEQVAFAKRENATDKTAYPM 309
QY 51 EYSEEGD-VOLNPIPPDFKLGSSGDEGRKVIYTONENIVYNSMYESFSISFWIRNK-- 107
DB 310 EYVLGNDKYSKNVIVDY-----GNQKGQQLIS-STNYINN--EDLSRMTVYVYNQPK 358
QY 108 -----WVSNLPGYTTIDSVKNNSGWSIGIISNPL-VFTLKONEDSEQSINFSYDINN 159
DB 359 KTYTKETFTVTLTGKFNPDAKNKIYEVTNQNFVDSFTPTDKLTVDTKFKITYSND 418
QY 160 APGYNKWFVTVTNMGMNKIYI-----NGKLIDTIKVK-----ELTG 198
DB 419 ----NKTATVLLNGQSSDKQYIIQQVAYPDNSSTNGKIDYILETQNGKSSWSNSYN 474
QY 199 INFSTIITFEINKIPDTGLITSDS-----DNINMWIRDFYIFAKELDGKDN----- 245
DB 475 VNGSSTANGDQKKYNLGDIYVWEDTKDGKODANEKIGKGVYVILKDSNGKELDRITTDEN 534
QY 246 -----ILFNSLQYTNVVKDYWGNDLRYNKE 270
DB 535 GKYQFTGLNGTYSVEFTLAGYPTTANAGTDDAVIDSDGLTTGVKID--ADNMILDSG 592
QY 271 YVMNIDYLNRYMYANSR-----QIVFNTRNNN-----D 300
DB 593 FYKTPKYSILGDYVWYDSNKGDKQDSTEKGIDKVKVTLNNEKGEVIGTKTDENGKYRFDN 652
QY 301 FNEG-YKIIIRKIRGNT-----NDRVRGGDI-----LYFDMTINNKAYNLFMKNET 346
DB 653 LDSGKYKVIKFEKPAGLTGTGTNTTEDDKADGGEVDVITDHDHDFLLDNGYY-----EET 708
QY 347 MYADNHSTED 356
DB 709 SDSDSDSDD 718

RESULT 10
US-09-815-242-5779
; Sequence 5779, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
```



Qy 300 -DFNEG-YKIIKIRNT-----NTRVRGGDILYFDMTINN-----KAYNLFMKNE 345  
Db 660 DNLDSGKYKVFIFKPAGLTQTNTTDDKADGGV--DVTITDHDFTLDNGYFEE 716  
Qy 346 TMYADNHSTED 356  
Db 717 TSDSDSDSDSD 727

RESULT 12  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 5.4%; Score 128; DB 10; Length 1430;  
Best Local Similarity 20.6%; Pred. No. 0.04;  
Matches 117; Conservative 73; Mismatches 188; Indels 190; Gaps 30;

Qy 10 YTNNSL-KDIINEYFNND--SKILSLQNRKNT---LVDFSGYNA-----EYSEEG 56  
Db 850 YESSALDSQLLYEGSFNFQDFYTKDSYTNKKIAQNVQLFKSWGVTSEMAPQYVSSD 909  
Qy 57 DVQNLPIPPDFKLGSGGDRGKVIYTONENIVNWSYSES-----ISFWIR 104  
Db 910 GSFLDSI-----IQNGYAFEDRYDLAKSKNNK--YGSQODMINAVKALHKSQIYIADW-- 961  
Qy 105 INKAVSNLPYTIIDSVK--NNSGWSIGIISNLFVTLTKQEDSE--QSINFSYDISNAPG 162  
Db 962 VPDQIYNLPGEVYVTRVNDYG-----EYKDKSEIANTLYAANKSNGKD 1007  
Qy 163 YNKFFVTVTNMNG-----NKIYINGKLID-----TIKVELTGNF----- 201  
Db 1008 YQARYGGAFLSALA KYPISFNTRQISNGKKIDPSEKITAWKAKYFNGTNILRGVGYVL 1067  
Qy 202 ---SKTITFEI-----NKIPDTGLI-----TSDSDNINMWIRD---F 232  
Db 1068 KDNASDKYFELKNGQTVLPQMTNKEASTGFVNDGNGMFPYFSGYQAKNSFVQDAKGNW 1127  
Qy 233 YIF-----AKELDGKJINILFNSLOY-----TNVVKDYGN--DLRYNKEYMYN 275  
Db 1128 YFDFNNGHMVYGLQALNGEVQYFSLNGVQVRESFLENADGSKNYFGLHGNRYSGYISFD 1187  
Qy 276 IDYLNRYMYANSROIV-FNTRRNNDP--NEGKIIKIRGNTNTRVRGGDILYFDMT 332  
Db 1188 NDSKRWYFDAGSVMAVLKTINGNTQYFDQDQGVQVGAWITGSDGKKR-----YFDDG 1240

Qy 333 INKAYNLFM--KNETMYADNHSTEDIYAGLRE-----QTKDINDN--- 372  
Db 1241 SGMAVNRFAANDKNGDWYLN--SDGIALVGVQTINGKTYTQGDKQIKGRIITDNKGL 1298  
Qy 373 -----IIFOIQPMNNTYYVASQ-----IFKSN-----F 395  
Db 1299 KYFLANGSELARNIFATDSQNNWYFSGDVAVTGSQTAGKKLYFASDGKQVKGSEFTY 1358  
Qy 396 NGE-----NISGICSGTYRFRGLGDMY 418  
Db 1359 NGKVHYVHADSGELQVNRFEADKGNWY 1386

RESULT 13  
US-10-047-676A-7  
; Sequence 7, Application US/10047676A  
; Patent No. US20020123105A1  
; GENERAL INFORMATION:  
; APPLICANT: Qi, Fengxia  
; APPLICANT: Caulfield, Page W.  
; APPLICANT: Chen, Ping W.  
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS  
; FILE REFERENCE: UAB-17403/22  
; CURRENT APPLICATION NUMBER: US/10/047,676A  
; CURRENT FILING DATE: 2002-03-21  
; PRIOR APPLICATION NUMBER: US 09/627,376  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 990  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-047-676A-7

Query Match 5.1%; Score 121.5; DB 12; Length 990;  
Best Local Similarity 21.9%; Pred. No. 0.087;  
Matches 107; Conservative 86; Mismatches 170; Indels 125; Gaps 29;

Qy 12 NNSLL-----KDIINEYFN--NIND--SKILSLQNRKNTLVDTSGY-----NAEVS- 53  
Db 254 NNSLELLEKLEIKAMITQDYTDNRIGSGNNSILALEKNKSOIVKANAYLRVDLYDHAEKL 313  
Qy 54 ---BEGDVQ--LNPFPDFKLGSGGEDRGKVIYTONENIVNWSYSESISFQWIRNKW 108  
Db 314 AQHTKSLQNLKVLSSF-----SSAVNSQKEIKNYHEKFIARYGYEQ-L-VPLQLLLNS- 366  
Qy 109 VSNL---PGYTIID-SYKNNSGWSIGIISNLFVTLTKQ-----NEDGEQSINFSY 154  
Db 367 TSGLGFPKGISQTEVSKQNNEDSKNOKIIEFLQKFEKALRDGKEILLSDDDLKDLNFTD 426  
Qy 155 DISNAPGYNKKFFVTY-----TNMMGNMKIYINGKLIDTI-----VKVELTGIN 200  
Db 427 EQQISGELYCFYFNFKSKKLEVSLSIGVSQLNGTFRFHSKLPNTIVTKNVNKTKEIFTEA 486  
Qy 201 FSKTITEINKIPDTG-----LITSDDNINMWIRDFYIFAKELDGKDNILFNSLOYTN 255  
Db 487 IPNTIITQLNEVPVFGGGMIMLSLKSQLELRN-YTTKKEMSINDIYVTRATSEEL-- 543  
Qy 256 VVKDYGNDRYRNEY-----YMNVIDYLNRYMYAN--SRQIVFENTERNNNDFN----- 302  
Db 544 ---YF-----YSKKYEKRVIFVMN---NMENYINGSKLLRELVESVNSDFQNTIPTFL 590  
Qy 303 ---BQYK-----IIIKIRGNTNTRVRGGDILYFDMTINNKAYNLFMK--NETMY 348  
Db 591 GSLDSYNHVPATYIKDIIIIKPEWTNRKSEAKTLDLKNWLTNNVPPVRMYKIDQIIY 650  
Qy 349 ADNHSTEDIYAI--GLRE-----QTKDI-----NDNIFIQ--IOPMNNNT-----YYASQ 389  
Db 651 LDLSTIDLMLFQSIKKHSFIQLLDVHSVCTNDTEILELWVPETRSVNAHQIYHAQN 710  
Qy 390 IFKSNFNG 397

Db 711 IYTLDSG 718

## RESULT 14

US-08-834-666A-20

; Sequence 20, Application US/08834666A

; Patent No. US20020044949A1

; GENERAL INFORMATION:

; APPLICANT: Kleanthous, Harold

; APPLICANT: Lissolo, Ling

; APPLICANT: Tomb, Jean-Francois

; APPLICANT: Miller, Charles

; APPLICANT: Al-Garawi, Amal

; TITLE OF INVENTION: 76 kDa Helicobacter Polypeptides and

; TITLE OF INVENTION: Corresponding Polynucleotide Molecules

; NUMBER OF SEQUENCES: 64

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Clark &amp; Elbing LLP

; STREET: 176 Federal Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/834,666A

; FILING DATE: 01-APR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,175

; REFERENCE/DOCKET NUMBER: 06132/038001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-428-0200

; TELEFAX: 617-428-7045

; TELEX:

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 483 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; FRAGMENT TYPE: internal

; FEATURE:

; NAME/KEY: Signal Sequence

; LOCATION: 1..16

; OTHER INFORMATION:

US-08-834-666A-20

Query Match 5.1%; Score 121; DB 8; Length 483;

Best Local Similarity 20.8%; Pred. No. 0.04;

Matches 96; Conservative 69; Mismatches 174; Indels 122; Gaps 22;

QY 12 NNSLLKDIINEYFNINDSKILSLQNRKNTLVDT-SGYNAEYSESGDVQLNFIFFDEKL 70

Db 38 NTGIEQVSNAY-ENLN-----NLLTRYNELKQTASNTSSTAQAI----- 78

QY 71 GSSGEGRGKIVTQNEINIVYNSMYESFISFWIRINKWYNSLPGVTIIDSVKNSGWSIG 130

Db 79 -NLKESARLKTTPNSANQAVSSALSSAVAMQVI--VSNLANNSLPTSEYN---KIN 130

QY 131 IISNPLVFTLKQEDSQSINFSYD-----ISNAPGYN-----KWFFVTYN 173

Db 131 AISOQLQNTL-ENKNDLKTENDYDHLLOASTIINTLOSQCPIGDGNGKPGWGINASGN 189

QY 174 --NMGMNKVIYNGKLDTIK----- 192

Db 190 ACNIEGTFNVAIT-SMDSAKKAADARRTAPESPQPSAFNADFNKNLNOVSSVINDT 248

QY 193 VKELTGINFSKTITTFEINKIPDT----GLITSDSDNINM---WIRDFYIFAKE----- 238

Db 249 ISYLKGDNLA-TIYNTLQKTPDSKGFQSLVSRSSYSYSLNETQYSEFQTTTKEFGHNFR 307

QY 239 ---LQKIDINILFNSLOYTVVVDYWGNDLRYNKREYMWVNDYLNRYMYANSROIVENT 294

Db 308 SVGLINSQSNNGAMGVQVLGYKQFFGKNKEFGTRYAF-PDYNHAYIKSN----FFNS 362

QY 295 RRNNDNFEGYKIIIRIRGTNDTR-----VRGGDILYFDMTINNKAYNLFMKNETMYA 349

Db 363 ASNVFTYAGASOLLNFINGSGDKNRKVSFGIFGGTALAGTTWLSQPMNLKTTNSAYSA 422

QY 350 --DNHSTEDIYALGLEQOKDINDNIIFOIQ-PMNNTYYIA 387

Db 423 KINNTNFQFLFTGLRLQ--GIHHGVELGVKIPITNTNYS 461

## RESULT 15

US-09-815-242-5809

; Sequence 5809, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-23

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5809

; LENGTH: 993

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-5809

Query Match 4.9%; Score 118; DB 10; Length 993;

Best Local Similarity 21.3%; Pred. No. 0.16;

Matches 104; Conservative 83; Mismatches 214; Indels 88; Gaps 23;

QY 5 FNFSYTNNSLLKDIINEYFNINDSKILSLQNRKNTL--VDTSGYNAEYSESGDVQLNP 62

Db 23 FNLFLNTEQIKYSDL--QLNAQMKESILTTFENLYCTLQEIINFDDGDKKVRDAKESLLKY 80

QY 63 IF-----PFDKLGSGDEGRGKVIVTQNEINIVYNSMYESFISFWIRINKWYNSLPGY-T 116

Db 81 LIRMSTRPTFGL-LSGINLGHFVNEPTRLKYGNSTQKVKYD-----GEMLYKLVSIE 134

QY 117 IIDSVKNSGWSIGITSNLFVFTLKO-----NEDSEOSINFSYDISNAPGYNKWEFF--T 170

Db 135 SIDEYQN----LKVWNSKAHIINDRIVLNEQSALYLNNDKTSFSIKNSLLYFIKTT 190

```

QY 171 VTNN--MMGNKTYING--KLIDTIKVKELTGINSKTIITFEINKIPDGLITSDDNIN 226
Db 191 VTNNITFSNLAEKINGEPEINDITKVYIHNLSVKEIISTIRPP-----LSYSDNIN 245
QY 227 MWIRDFYI---FAKELDGKDINILFNSLOYTNVVDYWGNDLRN-----KEY 271
Db 246 YILNKLSLHNDDEFVKTI--REIQKLIAYEKTEI--GFEELYKDIIHHMKALFKCKNY 300
QY 272 YMVN--IDYLNRYWY-----ANSRQIVENTRRNNNDENE-----GYKIIK 310
Db 301 LQIDTKIDMINNYLHQDIATNISEAYALLLSSNNIGFTDLKVLHNRFEKYGPEQLV- 359
QY 311 RIRGTNDTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHESTEDIYAIGLREOTKDIN 370
Db 360 NIKDLISDITGFGTTIFQEBETDCNNI--VNLKQKFLHALRNNDEIV--INEKDVESLIN 415
QY 371 DNIIFQIQPMNNTYYIASQIFKSNFNGENISGICSIGTYFRGLGGDWYRHNVLVPTV--- 427
Db 416 DNEINHYHAPMSADVYAEYLGRFYNOYNELIVPLTASFNAGATFGRFHLIDFETLA 475
QY 428 ----KQGN 432
Db 476 KLEHEKGHY 484

```

Search completed: November 7, 2002, 14:53:54  
Job time : 24 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:43:16 ; Search time 286 Seconds  
(without alignments)  
1014,440 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIPNFNFSYNNLSLKDI.....NYASLLESTTHGFPVSE 450

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

- 1: /cgn2\_6/ptodata/1/paa/pctus\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*
- 25: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*
- 26: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*
- 27: /cgn2\_6/ptodata/1/paa/US60\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2391	100.0	450	20	US-09-611-419A-10
2	2391	100.0	450	23	US-09-910-186A-10
3	2386	99.8	1291	1	PCT-US97-15394-60
4	2386	99.8	1291	11	US-08-704-159-60
5	2386	99.8	1291	13	US-08-954-302-1
6	2386	99.8	1291	21	US-09-791-537-13524

7	2386	99.8	1291	21	US-09-791-537-95055
8	2386	99.8	1291	26	US-10-205-516-6
9	2386	99.8	1291	26	US-10-271-012-60
10	2386	99.8	1301	26	US-10-205-516-20
11	2317	96.9	462	1	PCT-US97-15394-62
12	2317	96.9	462	11	US-08-704-159-62
13	2317	96.9	462	26	US-10-271-012-62
14	1824	76.3	1285	21	US-09-791-537-59796
15	1823	76.2	1285	21	US-09-791-537-16595
16	882.5	36.9	1280	21	US-09-791-537-75966
17	882.5	36.9	1280	21	US-09-791-537-112792
18	867.5	36.3	451	20	US-09-611-419A-12
19	867.5	36.3	1275	21	US-09-791-537-33117
20	867.5	36.3	1276	1	PCT-US97-15394-66
21	867.5	36.3	1276	11	US-08-704-159-66
22	867.5	36.3	1276	19	US-09-547-188-8
23	867.5	36.3	1276	21	US-09-791-537-13527
24	867.5	36.3	1276	21	US-09-791-537-92962
25	867.5	36.3	1276	26	US-10-205-516-8
26	867.5	36.3	1276	26	US-10-271-012-66
27	867.5	36.3	1286	26	US-10-205-516-22
28	864.5	36.2	451	23	US-09-910-186A-12
29	801.5	33.5	451	1	PCT-US97-15394-68
30	801.5	33.5	451	11	US-08-704-159-68
31	801.5	33.5	451	26	US-10-271-012-68
32	604.5	25.3	1297	1	PCT-US97-15394-77
33	604.5	25.3	1297	11	US-08-704-159-77
34	604.5	25.3	1297	19	US-09-547-188-14
35	604.5	25.3	1297	21	US-09-791-537-14255
36	604.5	25.3	1297	21	US-09-791-537-132350
37	604.5	25.3	1297	26	US-10-205-516-14
38	604.5	25.3	1297	26	US-10-271-012-77
39	604.5	25.3	1307	26	US-10-205-516-28
40	602.5	25.2	1252	26	US-10-205-516-10
41	602.5	25.2	1262	26	US-10-205-516-24
42	595.5	24.9	1250	1	PCT-US97-15394-50
43	595.5	24.9	1250	11	US-08-704-159-50
44	595.5	24.9	1250	26	US-10-271-012-50
45	594.5	24.9	449	23	US-09-910-186A-14

#### ALIGNMENTS

RESULT 1  
US-09-611-419A-10  
; Sequence 10, Application US/09611419A  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Leonard A.  
; APPLICANT: Byrne, Michael P.  
; APPLICANT: Middlebrook, John L.  
; APPLICANT: Lapenotiere, Hugh  
; APPLICANT: Clayton, Michael A.  
; APPLICANT: Brown, Douglas R.  
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
; FILE REFERENCE: A33626 067252.0105  
; CURRENT APPLICATION NUMBER: US/09/611,419A  
; PRIORITY FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: PCT/US00/12890  
; PRIOR FILING DATE: 2000-05-12  
; PRIOR APPLICATION NUMBER: 60/133,865  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,866  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,867  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,868  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,869  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/146,192  
; PRIOR FILING DATE: 1999-07-29

```
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Encoded polypeptide of a synthetic construct based
; OTHER INFORMATION: on Bontic HC
US-09-611-419A-10

Query Match      100.0%; Score 2391; DB 20; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.8e-212;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTIPFNFSYTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEYSEEGDVQL 60
DB 1 MTIPFNFSYTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEYSEEGDVQL 60
QY 61 NPIPPDFKLGSSGDRGKVIYVQENIVYNSMYESFSFWIRINKWYSNLPGYTIIDS 120
DB 61 NPIPPDFKLGSSGDRGKVIYVQENIVYNSMYESFSFWIRINKWYSNLPGYTIIDS 120
QY 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDINSNAPGYNKWFVYVTTNNMGNNK 180
DB 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDINSNAPGYNKWFVYVTTNNMGNNK 180
QY 181 IYINGKLIDITKVELTGINFSTIIFKINIPDTGLITSDSDINNMWIRDFYIFAKELD 240
DB 181 IYINGKLIDITKVELTGINFSTIIFKINIPDTGLITSDSDINNMWIRDFYIFAKELD 240
QY 241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYVMNIDYLNRYMYANSRQIVENTRRNND 300
DB 241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYVMNIDYLNRYMYANSRQIVENTRRNND 300
QY 301 FNEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360
DB 301 FNEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360
QY 361 GLRQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420
DB 361 GLRQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420
QY 421 NYLVPTVKQNGYASLLESTSTHMGFVPVSE 450
DB 421 NYLVPTVKQNGYASLLESTSTHMGFVPVSE 450
```

## RESULT 2

```
US-09-910-186A-10
; Sequence 10, Application US/09910186A
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research & Material Command
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; FILE REFERENCE: A33626-A 067252.0107
; CURRENT APPLICATION NUMBER: US/09/910,186A
; CURRENT FILING DATE: 2001-07-20
; PRIORITY APPLICATION NUMBER: PCT/US00/12890
; PRIORITY FILING DATE: 2000-05-12
; PRIORITY APPLICATION NUMBER: 09/611,419
; PRIORITY FILING DATE: 2000-07-06
; PRIORITY APPLICATION NUMBER: 60/133,865
; PRIORITY FILING DATE: 1999-05-12
; PRIORITY APPLICATION NUMBER: 60/133,866
; PRIORITY FILING DATE: 1999-05-12
; PRIORITY APPLICATION NUMBER: 60/133,867
; PRIORITY FILING DATE: 1999-05-12
; PRIORITY APPLICATION NUMBER: 60/133,868
; PRIORITY FILING DATE: 1999-05-12
; PRIORITY APPLICATION NUMBER: 60/133,869
; PRIORITY FILING DATE: 1999-05-12
; PRIORITY APPLICATION NUMBER: 60/133,873
```

```
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; OTHER INFORMATION: sequence
US-09-910-186A-10

Query Match      100.0%; Score 2391; DB 23; Length 450;
Best Local Similarity 100.0%; Pred. No. 2.8e-212;
Matches 450; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTIPFNFSYTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEYSEEGDVQL 60
DB 1 MTIPFNFSYTNNSLLKDIINEYFNINNSKILSLQNRKNTLVDTSGYNAEYSEEGDVQL 60
QY 61 NPIPPDFKLGSSGDRGKVIYVQENIVYNSMYESFSFWIRINKWYSNLPGYTIIDS 120
DB 61 NPIPPDFKLGSSGDRGKVIYVQENIVYNSMYESFSFWIRINKWYSNLPGYTIIDS 120
QY 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDINSNAPGYNKWFVYVTTNNMGNNK 180
DB 121 VKNNSGWSIGIISNLFVTLKQNEDESEQSINFSYDINSNAPGYNKWFVYVTTNNMGNNK 180
QY 181 IYINGKLIDITKVELTGINFSTIIFKINIPDTGLITSDSDINNMWIRDFYIFAKELD 240
DB 181 IYINGKLIDITKVELTGINFSTIIFKINIPDTGLITSDSDINNMWIRDFYIFAKELD 240
QY 241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYVMNIDYLNRYMYANSRQIVENTRRNND 300
DB 241 GKDNILFNSLQYTNVVKDYWGNDLRYNKYYVMNIDYLNRYMYANSRQIVENTRRNND 300
QY 301 FNEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360
DB 301 FNEGKIIIRKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360
QY 361 GLRQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420
DB 361 GLRQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSTGYRFLGGDWYRH 420
QY 421 NYLVPTVKQNGYASLLESTSTHMGFVPVSE 450
DB 421 NYLVPTVKQNGYASLLESTSTHMGFVPVSE 450
```

## RESULT 3

```
PCT-US97-15394-60
; Sequence 60, Application PC/TUS9715394
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: PCT/US97/15394  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPD-02304  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1291 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US97-15394-60

Query Match 99.8%; Score 2386; DB 1; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIFNIFSYTNNLSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEYSEEGDVQLN 61  
DB 843 TIFNIFSYTNNLSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEYSEEGDVQLN 902

QY 62 PIFPDEKLGSSGDEGRGVIVTQENIYVNSMYESFSISFWIRNKVWSNLPGYTIIDSV 121  
DB 903 PIFPDEKLGSSGDEGRGVIVTQENIYVNSMYESFSISFWIRNKVWSNLPGYTIIDSV 962

QY 122 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNAPGYNKWFVTVTNMMGNMKI 181  
DB 963 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNAPGYNKWFVTVTNMMGNMKI 1022

QY 182 YINGKLIDITIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRFYIFAKELDG 241  
DB 1023 YINGKLIDITIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRFYIFAKELDG 1082

QY 242 KDINILFNSLQYTNVVKDYGNDLRNKEYYVWIDYLNRYMYANSQIVFENTRRNNDF 301  
DB 1083 KDINILFNSLQYTNVVKDYGNDLRNKEYYVWIDYLNRYMYANSQIVFENTRRNNDF 1142

QY 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 361  
DB 1143 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202

QY 362 LREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSGTYRFLGSDWYRHN 421  
DB 1203 LREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSGTYRFLGSDWYRHN 1262

RESULT 4  
US-08-704-159-60  
Sequence 60, Application US/08704159  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
APPLICANT: Thalley, Bruce S.  
TITLE OF INVENTION: Multivalent Vaccine for Clostridium  
TITLE OF INVENTION: Botulinum Neurotoxin  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESS: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/704,159  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPD-02304  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 60:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1291 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-704-159-60

Query Match 99.8%; Score 2386; DB 1; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIFNIFSYTNNLSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEYSEEGDVQLN 61  
DB 843 TIFNIFSYTNNLSLLKDIINEYFNNINDSKILSLQNRKNTLVDTSGYNAEYSEEGDVQLN 902

QY 62 PIFPDEKLGSSGDEGRGVIVTQENIYVNSMYESFSISFWIRNKVWSNLPGYTIIDSV 121  
DB 903 PIFPDEKLGSSGDEGRGVIVTQENIYVNSMYESFSISFWIRNKVWSNLPGYTIIDSV 962

QY 122 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNAPGYNKWFVTVTNMMGNMKI 181  
DB 963 KNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNAPGYNKWFVTVTNMMGNMKI 1022

QY 182 YINGKLIDITIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRFYIFAKELDG 241  
DB 1023 YINGKLIDITIKVKELTGINFSTITFEINKIPDTGLITSDSDNINMWIRFYIFAKELDG 1082

QY 242 KDINILFNSLQYTNVVKDYGNDLRNKEYYVWIDYLNRYMYANSQIVFENTRRNNDF 301  
DB 1083 KDINILFNSLQYTNVVKDYGNDLRNKEYYVWIDYLNRYMYANSQIVFENTRRNNDF 1142

QY 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 361  
DB 1143 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIG 1202

QY 362 LREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSGTYRFLGSDWYRHN 421  
DB 1203 LREQTKDINDNIIFOQPMNNTYYASQIFKSNFNGENISGICSGTYRFLGSDWYRHN 1262

QY 422 YLVPVTKQGYASLLESTSTHWGFVPVSE 450  
DB 1263 YLVPVTKQGYASLLESTSTHWGFVPVSE 1291

RESULT 5  
US-08-954-302-1  
Sequence 1, Application US/08954302  
GENERAL INFORMATION:  
APPLICANT: Lance Simpson, Nikita Kiyatkin,  
APPLICANT: Andrew Maksymowich  
TITLE OF INVENTION: Compositions and Methods for Systemic  
TITLE OF INVENTION: Delivery of Oral Vaccines and Therapeutic Agents  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: Jane Massey Licata, Esq.  
STREET: 66 E. Main Street  
CITY: Marlton  
STATE: NJ



COUNTRY: USA  
ZIP: 08053  
COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM 486  
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/954,302  
FILING DATE: herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane Massey Licata  
REGISTRATION NUMBER: 32,257  
REFERENCE/DOCKET NUMBER: JEFF-0164  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (609) 779-2400  
TELEFAX: (609) 810-1454  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1291  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
US-08-954-302-1

Query Match 99.8%; Score 2386; DB 13; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIPNIFSYTNNLSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 61  
DB 843 TIPNIFSYTNNLSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 902

QY 62 PIPFPDKLGSSGEDRGKVIIVTONENIVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
DB 903 PIPFPDKLGSSGEDRGKVIIVTONENIVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962

QY 122 KNSGWSIGIISNLFVTLKQEDSEQSINFSYDINNNAPGYNKWFVTVTNMGNMKI 181  
DB 963 KNSGWSIGIISNLFVTLKQEDSEQSINFSYDINNNAPGYNKWFVTVTNMGNMKI 1022

QY 182 YINGKLIDTIKVKELTGINFSTITFEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 241  
DB 1023 YINGKLIDTIKVKELTGINFSTITFEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 1082

QY 242 KDINLFLNSLOYTNVVKDYWGNDLRNKEYVMYVNDYLNRYMYANSRQIVFTRNNND 301  
DB 1083 KDINLFLNSLOYTNVVKDYWGNDLRNKEYVMYVNDYLNRYMYANSRQIVFTRNNND 1142

QY 302 NEGKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 361  
DB 1143 NEGKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1202

QY 362 LREOTKDINDNIIFOQPMNNTYVYASQIFKSNFNGENISGICSGTFRFLGGDWYRHN 421  
DB 1203 LREOTKDINDNIIFOQPMNNTYVYASQIFKSNFNGENISGICSGTFRFLGGDWYRHN 1262

QY 422 YLVPVTKQGNYSALLESTSTHWGFPVSE 450  
DB 1263 YLVPVTKQGNYSALLESTSTHWGFPVSE 1291

RESULT 6  
US-09-791-537-13524  
; Sequence 13524, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB

; TITLE OF INVENTION: METHODS OF USE THEREOF  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 13524  
; LENGTH: 1291  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-09-791-537-13524

Query Match 99.8%; Score 2386; DB 21; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TIPNIFSYTNNLSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 61  
DB 843 TIPNIFSYTNNLSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNAEVEEGDVQLN 902

QY 62 PIPFPDKLGSSGEDRGKVIIVTONENIVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121  
DB 903 PIPFPDKLGSSGEDRGKVIIVTONENIVNSMYESFSISFWIRINKWVSNLPGYTIIDSV 962

QY 122 KNSGWSIGIISNLFVTLKQEDSEQSINFSYDINNNAPGYNKWFVTVTNMGNMKI 181  
DB 963 KNSGWSIGIISNLFVTLKQEDSEQSINFSYDINNNAPGYNKWFVTVTNMGNMKI 1022

QY 182 YINGKLIDTIKVKELTGINFSTITFEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 241  
DB 1023 YINGKLIDTIKVKELTGINFSTITFEINKIPDTGLITSDSNINMWIRDFYIFAKELDG 1082

QY 242 KDINLFLNSLOYTNVVKDYWGNDLRNKEYVMYVNDYLNRYMYANSRQIVFTRNNND 301  
DB 1083 KDINLFLNSLOYTNVVKDYWGNDLRNKEYVMYVNDYLNRYMYANSRQIVFTRNNND 1142

QY 302 NEGKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 361  
DB 1143 NEGKIIKIRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1202

QY 362 LREOTKDINDNIIFOQPMNNTYVYASQIFKSNFNGENISGICSGTFRFLGGDWYRHN 421  
DB 1203 LREOTKDINDNIIFOQPMNNTYVYASQIFKSNFNGENISGICSGTFRFLGGDWYRHN 1262

QY 422 YLVPVTKQGNYSALLESTSTHWGFPVSE 450  
DB 1263 YLVPVTKQGNYSALLESTSTHWGFPVSE 1291

RESULT 7  
US-09-791-537-95055  
; Sequence 95055, Application US/09791537  
; GENERAL INFORMATION:  
; APPLICANT: Bionomix, Inc.  
; APPLICANT: Debe, Derek  
; APPLICANT: Danzer, Joseph  
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME  
; FILE REFERENCE: 261/210  
; CURRENT APPLICATION NUMBER: US/09/791,537  
; CURRENT FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 153055  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 95055  
; LENGTH: 1291  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum phage  
US-09-791-537-95055

Query Match 99.8%; Score 2386; DB 21; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 TTPNFNFSYTNSSLLKDIINEYFNNDKSLQNRKNTLVDTSQYNAEYSEEGDVLN 61
DB 843 TTPNFNFSYTNSSLLKDIINEYFNNDKSLQNRKNTLVDTSQYNAEYSEEGDVLN 902
QY 62 PTFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 121
DB 903 PTFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 962
QY 122 KNSGWSIGIISNLFVTLKQNEDEQISYDINSNAPGNKWFVTVTNMNGNMI 181
DB 963 KNSGWSIGIISNLFVTLKQNEDEQISYDINSNAPGNKWFVTVTNMNGNMI 1022
QY 182 YINGKLIDTIKVKELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241
DB 1023 YINGKLIDTIKVKELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082
QY 242 KDINILFNSLQYTNVVKDYWGNDLRYNKKEYYVWIDYLNRYMYANSQIVFNTRRNND 301
DB 1083 KDINILFNSLQYTNVVKDYWGNDLRYNKKEYYVWIDYLNRYMYANSQIVFNTRRNND 1142
QY 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAFLFKMNETYADNHSTEDIYAIG 361
DB 1143 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAFLFKMNETYADNHSTEDIYAIG 1202
QY 362 LREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 421
DB 1203 LREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 1262
QY 422 YLVPVTKQGNVYASLLESTSTHGWFPVSE 450
DB 1263 YLVPVTKQGNVYASLLESTSTHGWFPVSE 1291

RESULT 8
US-10-205-516-6
; Sequence 6, Application US/10205516
; GENERAL INFORMATION:
; APPLICANT: Zhong, Jun
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
; FILE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
; CURRENT REFERENCE: J2bt1
; CURRENT APPLICATION NUMBER: US/10/205,516
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1291
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-205-516-6

Query Match 99.8%; Score 2386; DB 26; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.9e-211;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTPNFNFSYTNSSLLKDIINEYFNNDKSLQNRKNTLVDTSQYNAEYSEEGDVLN 61
DB 843 TTPNFNFSYTNSSLLKDIINEYFNNDKSLQNRKNTLVDTSQYNAEYSEEGDVLN 902
QY 62 PTFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 121
DB 903 PTFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 962
QY 122 KNSGWSIGIISNLFVTLKQNEDEQISYDINSNAPGNKWFVTVTNMNGNMI 181
DB 963 KNSGWSIGIISNLFVTLKQNEDEQISYDINSNAPGNKWFVTVTNMNGNMI 1022
QY 182 YINGKLIDTIKVKELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241
DB 1023 YINGKLIDTIKVKELTGINFSTIIFEINKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1082
QY 242 KDINILFNSLQYTNVVKDYWGNDLRYNKKEYYVWIDYLNRYMYANSQIVFNTRRNND 301

```

```

DB 1083 KDINILFNSLQYTNVVKDYWGNDLRYNKKEYYVWIDYLNRYMYANSQIVFNTRRNND 1142
QY 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAFLFKMNETYADNHSTEDIYAIG 361
DB 1143 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAFLFKMNETYADNHSTEDIYAIG 1202
QY 362 LREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 421
DB 1203 LREQTKDINDNIIFOIQPMNNTYYASQIFKSNFNGENISGICISGTYRFLGSDWYRHN 1262
QY 422 YLVPVTKQGNVYASLLESTSTHGWFPVSE 450
DB 1263 YLVPVTKQGNVYASLLESTSTHGWFPVSE 1291

RESULT 9
US-10-271-012-60
; Sequence 60, Application US/10271012
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/271,012
; FILING DATE: 15-Oct-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/704,159
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1291 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 60:
US-10-271-012-60

Query Match 99.8%; Score 2386; DB 26; Length 1291;
Best Local Similarity 100.0%; Pred. No. 3.9e-211;
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTPNFNFSYTNSSLLKDIINEYFNNDKSLQNRKNTLVDTSQYNAEYSEEGDVLN 61
DB 843 TTPNFNFSYTNSSLLKDIINEYFNNDKSLQNRKNTLVDTSQYNAEYSEEGDVLN 902
QY 62 PTFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 121
DB 903 PTFPFDFKLGSGEDRGKVIYQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDS 962
QY 122 KNSGWSIGIISNLFVTLKQNEDEQISYDINSNAPGNKWFVTVTNMNGNMI 181

```

Db 963 KNSGWSGIISNLFVTLKQNEDESGSINFSYDISNAPGYNKFFVTVTNMGMNMI 1022  
Qy 182 YNGKLIDITIKVKELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 241  
Db 1023 YNGKLIDITIKVKELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 1082  
Qy 242 KDINILFNSLOYTNVVDYDGNLDLYNKEYYMWYVNDYLNRYMYANSRQIVFTRNNNDP 301  
Db 1083 KDINILFNSLOYTNVVDYDGNLDLYNKEYYMWYVNDYLNRYMYANSRQIVFTRNNNDP 1142  
Qy 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIG 361  
Db 1143 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIG 1202  
Qy 362 LREQTKDINDNIIFOIPMNNTYYASQIFKSNFNGENISGICSGITYRFLGGDWYRHN 421  
Db 1203 LREQTKDINDNIIFOIPMNNTYYASQIFKSNFNGENISGICSGITYRFLGGDWYRHN 1262  
Qy 422 YLPTVKOGNYASLLESTSTHWGFPVSE 450  
Db 1263 YLPTVKOGNYASLLESTSTHWGFPVSE 1291  
RESULT 10  
US-10-205-516-20  
; Sequence 20, Application US/10205516  
; GENERAL INFORMATION:  
; APPLICANT: Zhong, Jun  
; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
; FILE REFERENCE: Jb7x1  
; CURRENT APPLICATION NUMBER: US/10/205,516  
; CURRENT FILING DATE: 2002-07-25  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 1301  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-10-205-516-20  
Query Match 99.8%; Score 2386; DB 26; Length 1301;  
Best Local Similarity 100.0%; Pred. No. 3.9e-211;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 TIFPNIFSYTNNSLLKDIIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61  
Db 847 TIFPNIFSYTNNSLLKDIIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGVDQLN 906  
Qy 62 PIFPDEFKLGSGEDRGKVIYQENIVYNSMYESFSISFWIRINKWWSNLPYTIIDSV 121  
Db 907 PIFPDEFKLGSGEDRGKVIYQENIVYNSMYESFSISFWIRINKWWSNLPYTIIDSV 966  
Qy 122 KNSGWSGIISNLFVTLKQNEDESGSINFSYDISNAPGYNKFFVTVTNMGMNMI 181  
Db 967 KNSGWSGIISNLFVTLKQNEDESGSINFSYDISNAPGYNKFFVTVTNMGMNMI 1026  
Qy 182 YNGKLIDITIKVKELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 241  
Db 1027 YNGKLIDITIKVKELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDG 1086  
Qy 242 KDINILFNSLOYTNVVDYDGNLDLYNKEYYMWYVNDYLNRYMYANSRQIVFTRNNNDP 301  
Db 1087 KDINILFNSLOYTNVVDYDGNLDLYNKEYYMWYVNDYLNRYMYANSRQIVFTRNNNDP 1146  
Qy 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIG 361  
Db 1147 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIG 1206  
Qy 362 LREQTKDINDNIIFOIPMNNTYYASQIFKSNFNGENISGICSGITYRFLGGDWYRHN 421  
Db 1207 LREQTKDINDNIIFOIPMNNTYYASQIFKSNFNGENISGICSGITYRFLGGDWYRHN 1266

Qy 422 YLPTVKOGNYASLLESTSTHWGFPVSE 450  
Db 1267 YLPTVKOGNYASLLESTSTHWGFPVSE 1295  
RESULT 11  
PCT-US97-15394-62  
; Sequence 62, Application PC/TUS9715394  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; TITLE OF INVENTION: Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US97/15394  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
PCT-US97-15394-62  
Query Match 96.9%; Score 2317; DB 1; Length 462;  
Best Local Similarity 99.8%; Pred. No. 2.1e-205;  
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 14 SLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGVDQLNFIFFDFKLGSS 73  
Db 26 ALLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGVDQLNFIFFDFKLGSS 85  
Qy 74 GEDRGKVIYQENIVYNSMYESFSISFWIRINKWWSNLPYTIIDSVNNSGWSGIIS 133  
Db 86 GEDRGKVIYQENIVYNSMYESFSISFWIRINKWWSNLPYTIIDSVNNSGWSGIIS 145  
Qy 134 NFLVETLKQNEDESGSINFSYDISNAPGYNKFFVTVTNMGMNMIYINGKLIDITIKV 193  
Db 146 NFLVETLKQNEDESGSINFSYDISNAPGYNKFFVTVTNMGMNMIYINGKLIDITIKV 205  
Qy 194 KELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDGKDINILFNSLOY 253  
Db 206 KELTGINFSTKITTEINKIPDGLTSDSDINNMWIRDFYIFAKELDGKDINILFNSLOY 265  
Qy 254 TNVVDYDGNLDLYNKEYYMWYVNDYLNRYMYANSRQIVFTRNNNDPNEGKIIIRK 313  
Db 266 TNVVDYDGNLDLYNKEYYMWYVNDYLNRYMYANSRQIVFTRNNNDPNEGKIIIRK 325  
Qy 314 GNTNDRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIGLREQTKDINDN 373  
Db 326 GNTNDRVRGGDILYFDMTINNKAYNLFPMKNETMYADNHSTEDIYAIGLREQTKDINDN 385

QY 374 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYRFRGLGGDWYRHNLYLPTVKQGNYA 433  
DB 386 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYRFRGLGGDWYRHNLYLPTVKQGNYA 445  
QY 434 SLESTSTHWGFVPVSE 450  
DB 446 SLESTSTHWGFVPVSE 462

RESULT 12  
US-08-704-159-62  
; Sequence 62, Application US/08704159  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Botulinum Vaccine For Clostridium  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08704,159  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-704-159-62

Query Match 96.9%; Score 2317; DB 11; Length 462;  
Best Local Similarity 99.8%; Pred. No. 2.1e-205;  
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLKLDIINEYFNNDKSKLSLQNKNTLVDTSGYNAEVSEEGVQLNPFPDFKLGSS 73  
DB 26 ALLKLDIINEYFNNDKSKLSLQNKNTLVDTSGYNAEVSEEGVQLNPFPDFKLGSS 85

QY 74 GEDRGKVIYTONENIVYNSMYESFSISFWIRKWSNLPCTIIDSVKNSGWSIGIIS 133  
DB 86 GEDRGKVIYTONENIVYNSMYESFSISFWIRKWSNLPCTIIDSVKNSGWSIGIIS 145

QY 134 NFIIVFTLKQNEDESEGSINFSYDISNNAFGYKWFVTVTNMGMNMIYINGKLIDTIKV 193  
DB 146 NFIIVFTLKQNEDESEGSINFSYDISNNAFGYKWFVTVTNMGMNMIYINGKLIDTIKV 205

QY 194 KLTGTFNFKTITFEINKIPDTGLTSSDNIINMIRFPIFAKELDGKDNIILFNSLQY 253  
DB 206 KLTGTFNFKTITFEINKIPDTGLTSSDNIINMIRFPIFAKELDGKDNIILFNSLQY 265

QY 254 TNNVKDYWGNDLRNKEYYVWVIDYLNRYMYANSRQIVFNTRRNNDFNEGYKIIIKRIR 313  
DB 266 TNNVKDYWGNDLRNKEYYVWVIDYLNRYMYANSRQIVFNTRRNNDFNEGYKIIIKRIR 325

QY 314 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLEQTKDINDNI 373  
DB 326 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGLEQTKDINDNI 385  
QY 374 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYRFRGLGGDWYRHNLYLPTVKQGNYA 433  
DB 386 IFQIQPMNNTYYASQIFKSNFNGENISGICISGTYRFRGLGGDWYRHNLYLPTVKQGNYA 445  
QY 434 SLESTSTHWGFVPVSE 450  
DB 446 SLESTSTHWGFVPVSE 462

RESULT 13  
US-10-271-012-62  
; Sequence 62, Application US/10271012  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/271,012  
; FILING DATE: 15-Oct-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/704,159  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 62:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 62:  
US-10-271-012-62

Query Match 96.9%; Score 2317; DB 26; Length 462;  
Best Local Similarity 99.8%; Pred. No. 2.1e-205;  
Matches 436; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 14 SLKLDIINEYFNNDKSKLSLQNKNTLVDTSGYNAEVSEEGVQLNPFPDFKLGSS 73  
DB 26 ALLKLDIINEYFNNDKSKLSLQNKNTLVDTSGYNAEVSEEGVQLNPFPDFKLGSS 85

QY 74 GEDRGKVIYTONENIVYNSMYESFSISFWIRKWSNLPCTIIDSVKNSGWSIGIIS 133  
DB 86 GEDRGKVIYTONENIVYNSMYESFSISFWIRKWSNLPCTIIDSVKNSGWSIGIIS 145

QY 134 NFIIVFTLKQNEDESEGSINFSYDISNNAFGYKWFVTVTNMGMNMIYINGKLIDTIKV 193  
DB 146 NFIIVFTLKQNEDESEGSINFSYDISNNAFGYKWFVTVTNMGMNMIYINGKLIDTIKV 205

```
QY 194 KELTGINFSTKTIPEINIKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLOY 253
Db 206 KELTGINFSTKTIPEINIKIPDTGLITSDSDNINMWIRDFYIFAKELDGKDINILFNSLOY 265
QY 254 TNVVDYWGNDLRNKEYYVWVNDYLNRYMYANSRQIVFNTNRNNDNFEGYKIIKRI 313
Db 266 TNVVDYWGNDLRNKEYYVWVNDYLNRYMYANSRQIVFNTNRNNDNFEGYKIIKRI 325
QY 314 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGREQTKDINDMI 373
Db 326 GNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAIGREQTKDINDMI 385
QY 374 IFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHNLYLVPVKGNYA 433
Db 386 IFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHNLYLVPVKGNYA 445
QY 434 SLESTSTHWGFVPVSE 450
Db 446 SLESTSTHWGFVPVSE 462

RESULT 14
US-09-791-537-59796
; Sequence 59796, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomimix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMB
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59796
; LENGTH: 1285
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-09-791-537-59796

Query Match 76.3%; Score 1824; DB 21; Length 1285;
Best Local Similarity 77.9%; Pred. No. 5e-159;
Matches 352; Conservative 36; Mismatches 56; Indels 8; Gaps 4;

QY 2 TIPNIFSYTNNLSLLKDIINEYFNNDKILSLQNKNTLVDTSYNAEVSSEGDVQLN 61
Db 839 TIPNIFSYTNNLSLLKDIINEYFNNDKILSLQNKNTLVDTSYNAEVSSEGDVQLN 898
QY 62 PIFPDEFKLGSSGDRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121
Db 899 PIFPDEFKLGSSGDRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 958
QY 122 KNSGWSIGIISNLFVTLKQNEDESEQINFSYDISKNAAGYKNKWFVTVTNNMGNMKI 181
Db 959 KNSGWSIGIISNLFVTLKQNEDESEQINFSYDISKNAAGYKNKWFVTVTNNMGNMKI 1018
QY 182 YINGKLIDTIKVKELTGINFSTKTIPEINIKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDTIKVKELTGINFSTKTIPEINIKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1078
QY 242 KDINILFNSLOYTNVVDYWGNDLRNKEYYVWVNDYLNRYMYANSRQIVFNTNRNNDNF 301
Db 1079 KDINILFNSLOYTNVVDYWGNDLRNKEYYVWVNDYLNRYMYANSRQIVFNTNRNNDNF 1138
QY 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 360
Db 1139 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 1193
QY 361 GLREQTKD-INDNIIFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGD-WY 418
Db 1194 GALTQPMDEIRKYSFIQPCNTFDYASQIFLSSNATNRLGILSIGSYFSLGDDYWF 1253
```

Search completed: November 7, 2002, 14:52:41  
Job time : 294 secs

```
QY 419 RHNLYLVPVTKQGNVYASLLESTSTHWGFVPVSE 450
Db 1254 NHEYLPVVKIEHYASLLESTSTHWGFVPVSE 1285

RESULT 15
US-09-791-537-16595
; Sequence 16595, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomimix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY ME
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16595
; LENGTH: 1285
; TYPE: PRT
; ORGANISM: Clostridium botulinum phase d
US-09-791-537-16595

Query Match 76.2%; Score 1823; DB 21; Length 1285;
Best Local Similarity 77.7%; Pred. No. 6.1e-159;
Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;

QY 2 TIPNIFSYTNNLSLLKDIINEYFNNDKILSLQNKNTLVDTSYNAEVSSEGDVQLN 61
Db 839 TIPNIFSYTNNLSLLKDIINEYFNNDKILSLQNKNTLVDTSYNAEVSSEGDVQLN 898
QY 62 PIFPDEFKLGSSGDRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 121
Db 899 PIFPDEFKLGSSGDRGVIVTQENIVYNSMYESFSISFWIRINKWVSNLPGYTIIDSV 958
QY 122 KNSGWSIGIISNLFVTLKQNEDESEQINFSYDISKNAAGYKNKWFVTVTNNMGNMKI 181
Db 959 KNSGWSIGIISNLFVTLKQNEDESEQINFSYDISKNAAGYKNKWFVTVTNNMGNMKI 1018
QY 182 YINGKLIDTIKVKELTGINFSTKTIPEINIKIPDTGLITSDSDNINMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDTIKVKELTGINFSTKTIPEINIKIPDTGLITSDSDNINMWIRDFYIFAKELDG 1078
QY 242 KDINILFNSLOYTNVVDYWGNDLRNKEYYVWVNDYLNRYMYANSRQIVFNTNRNNDNF 301
Db 1079 KDINILFNSLOYTNVVDYWGNDLRNKEYYVWVNDYLNRYMYANSRQIVFNTNRNNDNF 1138
QY 302 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 360
Db 1139 NEGKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMYADNHST-EDIYAI 1193
QY 361 GLREQTKD-INDNIIFQOPMNTYVYASQIFKSNFNGENISGICSTGYRFRLLGGD-WY 418
Db 1194 GALTQPMDEIRKYSFIQPCNTFDYASQIFLSSNATNRLGILSIGSYFSLGDDYWF 1253
QY 419 RHNLYLVPVTKQGNVYASLLESTSTHWGFVPVSE 450
Db 1254 NHEYLPVVKIEHYASLLESTSTHWGFVPVSE 1285
```

GenCore version 5.1.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:44:06 ; Search time 15 Seconds  
(without alignments)  
212.562 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391  
Sequence: 1 MTIPFNIFSYTNLSLKDI.....NYASLLESTSHWGFVPVSE 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 28998 seqs, 7085409 residues

Total number of hits satisfying chosen parameters: 28998

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/1/paa/PTCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	122	5.1	912	6	US-10-092-411A-2993
2	121.5	5.1	1335	6	US-10-092-411A-3716
3	119.5	5.0	930	6	US-10-092-411A-5314
4	117	4.9	691	6	US-10-092-411A-4675
5	116.5	4.9	1151	6	US-10-092-411A-3242
6	108	4.5	976	1	PCT-US02-05068-98
7	107.5	4.5	897	5	US-09-721-456-189
8	107.5	4.5	1007	5	US-09-721-456-187
9	107.5	4.5	1031	5	US-09-721-456-110
10	106	4.4	1010	6	US-10-092-411A-5178
11	105	4.4	10182	6	US-10-092-411A-3159
12	104.5	4.4	497	6	US-10-092-411A-5114
13	104.5	4.4	695	6	US-10-092-411A-4341
14	102	4.3	3696	6	US-10-092-411A-5080
15	100	4.2	527	6	US-10-092-411A-4465
16	100	4.2	804	6	US-10-092-411A-5218
17	99.5	4.2	670	6	US-10-092-411A-2940
18	99.5	4.2	1041	6	US-10-145-087A-498
19	99.5	4.2	1041	6	US-10-143-031A-498
20	99.5	4.2	1041	6	US-10-145-092A-498
21	99.5	4.2	1041	6	US-10-162-522A-498
22	99.5	4.2	1041	6	US-10-165-038A-498
23	99.5	4.2	1041	6	US-10-165-353-498
24	99.5	4.2	1041	6	US-10-170-481A-498
25	99.5	4.2	1041	6	US-10-172-039A-498
26	99.5	4.2	1041	6	US-10-145-016A-498

27	99.5	4.2	1041	6	US-10-145-088A-498	Sequence 498, App
28	99.5	4.2	1041	6	US-10-145-129A-498	Sequence 498, App
29	99.5	4.2	1041	6	US-10-165-353A-498	Sequence 498, App
30	99.5	4.2	4544	6	US-10-085-198-22	Sequence 22, Appl
31	99	4.1	956	6	US-10-092-411A-4452	Sequence 4452, Ap
32	98.5	4.1	422	6	US-10-092-411A-5230	Sequence 5230, Ap
33	98.5	4.1	806	6	US-10-092-411A-4314	Sequence 4314, Ap
34	97.5	4.1	428	6	US-10-092-411A-2942	Sequence 2942, Ap
35	97	4.1	509	6	US-10-092-411A-2996	Sequence 2996, Ap
36	96.5	4.0	430	6	US-10-092-411A-4712	Sequence 4712, Ap
37	96	4.0	1162	6	US-10-092-411A-4008	Sequence 4008, Ap
38	95	4.0	331	6	US-10-092-411A-3626	Sequence 3626, Ap
39	95	4.0	1151	6	US-10-060-019-31	Sequence 31, Appl
40	94.5	4.0	366	6	US-10-092-411A-4412	Sequence 4412, Ap
41	93.5	3.9	840	5	US-09-721-456-190	Sequence 190, App
42	93.5	3.9	872	5	US-09-721-456-221	Sequence 221, App
43	93.5	3.9	1183	6	US-10-092-411A-3530	Sequence 3530, App
44	93	3.9	267	6	US-10-092-411A-3520	Sequence 3520, Ap
45	93	3.9	478	6	US-10-271-145-2	Sequence 2, Appl

## ALIGNMENTS

### RESULT 1

US-10-092-411A-2993  
; Sequence 2993, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 2993  
; LENGTH: 912  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-2993

Query Match	5.1%	Score 122;	DB 6;	Length 912;
Best Local Similarity	20.6%	Pred. No. 0.26;		
Matches	96;	Conservative	72;	Mismatches 154; Indels 144; Gaps 23;
QY	3	IPFNIFSYTN-----SLKDIINEYFNINDSKILSLQNRKNTLVDTSGYNA--E	51	
DB	374	VFFDQMPFTASLKGHPNINITYLLOCI--EYKG--REYELLVRKIQKNTLANKKVYSKEE	429	
QY	52	VSEEDVOLNIFPDEFKLGSGEDRGKVIYQTNENIVY-----NSMYESFSISFWIRINK	107	
DB	430	IFQYGVV-----NELIDYNNLYKHPNREIHSFGDKYLYENE	470	
QY	108	WVSNLPGTIIDSVKNNSGWIGIISNLFVTL-----KQED-----SEQSINF	152	
DB	471	QSI-----MSIISKILLSNESVEGYSNVEFWLNNETSLDCKREKKEILLRMFSNISKISM	526	
QY	153	SYDISNNAFGYNKWFVVTNNMGMNKIYI---NGKLIDTIKVKELTGINFSTITFEI	209	
DB	527	VY-----GAAGTGSKLINHCNFFYDKDVIVANTTAVDNIKK-----IKLSNIKTSTI	578	
QY	210	NKIPDTGLTSDSDMINWIRDFYIFAKELGDKDINILFNSLQYTNVVKYWGNDLYNK	269	
DB	579	SK-----FLYNDKEKYDLLIID---EAGTVSNKDMNRILENKOPELL-----LIVGD	622	
QY	270	EYMYNIDVLYNMYVAN---SRQIVENTRRNNDFNEGYKIIKRIKRGNTNDTRVRGGDI	326	

Db 623 NYQIESIDFGNWFIAKDVLSKNII-----NELTDMY-----RTKNDL 661

Qy 327 LYFDMTINKAYNLFMKNFYADNHST-----EDYVAGLRE- 364

Db 662 LYFKSVREKSNII---NELINNKYSTRLDESIFNEFNKDEILCLNDYDGIYGINNR 718

Qy 365 --QKNDINDNIIFOIQP-----MNTVYASQIFKSNFNGENI 400

Db 719 LQANKNDVINGVKEYKVGDPILFNETKYSPILF-NNLKSII 763

RESULT 2

US-10-092-411A-3716

; Sequence 3716, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092.411A

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676

; SEQ ID NO 3716

; LENGTH: 1335

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-092-411A-3716

Query Match 5.1%; Score 121.5; DB 6; Length 1335;

Best Local Similarity 19.8%; Pred. No. 0.46;

Matches 97; Conservative 74; Mismatches 193; Indels 127; Gaps 23;

Qy 10 YTNSSLKDIINEYFNININDSKILSLQNRKNTLVDSGYNAEVSSEGDVQLNPIFFDPK 69

Db 900 YANRTEL---INKQRTGNTYVLLQDGTSTPL-----GWNINDVITQN-- 942

Qy 70 LGSSGDEGRKIVYQNEINIVNSMY-----ESFSIFWTRINKWSNLP 114

Db 943 IGKQTSIGKYSVFTNGLYSIANGTKNOQLAPNTLANQAFNASKAYVGK-----D 996

Qy 115 YTIIDSVKNSGWSIGIISNPLVFTLTKQNSDEQSINFSYD-ISNAPGYNKWFFVTVTN 173

Db 997 LYLYGTVNNRTGW---IAAKDLI---QNSTDAQSTPYNTYFVINNSKSY---FYMDPTK 1046

Qy 174 NMGMNMTIYINGKLIDITKYKELGINF-----SKTITEINKIPDGLIT 219

Db 1047 ANYSKLKPYE-QFTYIKQKNINGVWYQGLLDGKYVWIKSTDIVKIKYAYTGWTL 1105

Qy 220 SDDSNIN-----MWIRDFYIFAKELDGKIDINILFN--SLOYTNVVKDYGN 263

Db 1106 NNAINIQLAKYKQVQNEPLKWSNANSYQIK--NAMDTRKLANDSLKY-----Q 1154

Qy 264 DLRYNKEYMYNIDYLAHYM---YANSRQIVENTRNNNDNFEGYKIIKIR-CONTND 318

Db 1155 FURLDPOYL-SQAALNKLKGVLENQGAQAFSQAARKYGLNEIYLISHALVETGNGTS 1213

Qy 319 TVRGGDIL--YFDMTINNKAYNLFMKNFYADNHSTEDYVAGLREQTKDINDNIIFQ 376

Db 1214 QLAGGDVSKGFTTKTKGHYHVGIGA---FDNNALVDGIRYAKNAGWTSVSKAIIIG 1270

Qy 377 IQPMNTYVYASQ--IFKSNFNGENISGICSGTYRFLGGDWYRN-----421

Db 1271 AKFIGNSVKAGNTLYKMRWNPN-----PGTHQYATDINWANNVNAQVLYKOFYDKIGE 1324

Qy 422 ----YLVPTVK 428

Db 1325 VGKFEIPTYK 1335

RESULT 3

US-10-092-411A-5314

; Sequence 5314, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092.411A

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676

; SEQ ID NO 5314

; LENGTH: 930

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-10-092-411A-5314

Query Match 5.0%; Score 119.5; DB 6; Length 930;

Best Local Similarity 21.0%; Pred. No. 0.39;

Matches 98; Conservative 63; Mismatches 184; Indels 121; Gaps 23;

Qy 9 SYTN-----NSLLKDIINEYFNININDSKILSLQ-NRKNTLVDSGYNAEVSSEGDVQ 59

Db 224 SYKNIDEKISNQDELNLPINEVKNVRPLSTTSAPSSKRVTV-----NQLAAEGGSNV 278

Qy 60 LNPIFFDPFKLGGSGEDGRKIVYQNEINIVNSMYESFSIFWIRINKWSNLPGYTII 118

Db 279 NHLIKVTDQSITTEGYDSDGIKAHAENLIYDVTFE-----VD 317

Qy 119 DSVKNSGWSIGIISN-----FLVFTLKQNSDEQSINFSYDISNAPGYNKWFFV 169

Db 318 DKVSGDTMTVIDKNTVPSDLTDSFAIPKIKDN-SGEIATGYDNTNKKQTYFTFDIV 376

Qy 170 TVTNMNMGNMKI--YI-----NGKLIDITKYKELTGINFSTIITEINKIPD--TGLI 218

Db 377 DKYENIRAKHLKLSYIDKSKVPNNNTKLDVEYKLTSSVN--KITIVYQKPNENRTANL 434

Qy 219 TSDSDNIN-----MWIRDFYIFAKELDGKIDINILFNLSLOYTNVVKDYGNLRYNK 269

Db 435 QSMFTNIDTKNHTVEQTYIINPLRYSAKE---TNVAISGNGDEGSTIIDD---STIKVYK 489

Qy 270 EYVMVNTDYLNR-YMVANSRQIVENTRNNNDNFEGYKIIKIRGNTNDTRVRGGDI-- 326

Db 490 VGDQNLPDSNRIYDIYSEYDV-----TNDYQAQ-----LGNNDVNINFGNIDS 534

Qy 327 LYFDMTI-----NKNAYNLFMKNFYADNHSTEDYVAGLREQTKDINDNIIFOIQ-- 378

Db 535 PYLIKVSKYDPKNDDYTTIQQTVMTQTTINEYTGFF-----RTASYDNTIATSTSSG 587

Qy 379 -----PMNNTYVYASQIFK-----SNFNGENISGICSGTY 409

Db 588 QCGDLPEKTYKIGDYWEDVDKDGQIONDNEKPLSNVLTLY 633

RESULT 4

US-10-092-411A-4675

; Sequence 4675, Application US/10092411A

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-101

; CURRENT APPLICATION NUMBER: US/10/092.411A

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: US 09/134,001

; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 4675  
; LENGTH: 691  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-4675

Query Match 4.9%; Score 117; DB 6; Length 691;  
Best Local Similarity 19.3%; Pred. No. 0.4;  
Matches 116; Conservative 84; Mismatches 176; Indels 224; Gaps 33;  
Db 1 MTIPNFISYTNNS-----LLKDIINEYFNINDSKIL----- 33  
5 MSIPNL--PTNSTMINELCTLOSRTINIKGEVLITELYDDYFFKNDWHITAFNKQF 62  
34 --SLQN-----RKNLT--VDTSGYNAE-----VSEE-----GDVQLNPIPPFD 67  
63 QDSIKYRDKRNVFFRIKSKNLNLEFKYLFKLIVKEDWSLNFNTGAVKLNKIAK- 121  
68 FKLGSGEDRGKVIYQENIYVNSWYSEFSISFWIRN-----KWV--SNLP-----G 114  
122 -----FNEVYNLSLLDCLDINTLEKWFNWLNNPIKERS 159  
115 YTIIDSVKNNSGSIGITISNLFVFTLKQNEDESEQINSFYDISN---NAPGYNKWFVTV 171  
160 STIVFGDYEVKGLASFLKMYINLIKIDKREWEKDKWDIRNLEKYGLSYNK-----TL 215  
172 TNNMGNMAYIKGLIDITKVKELTGINFESKITIFEINKIPDTGLITSDSDNNMWRD 231  
216 TGN-----YLNFEKIESIKRELA-----KKILKNLITGD---TAFATAR 253  
232 FYI-----FAKELDKDINILFNSLQ-----YTNVVKYDWGN--DLRYNKEYYMWNI 278  
254 FYIRVLTREFQNFISKNKTRNSNELDRCHIEAYTEFLFEYAANKHLQSTKNFVREELKT 313  
279 LNYM-----YANSQIVF-----NTRNNNDNEGYKIIIKIRGTND-- 318  
314 IRRFLNDITQNAIAPYQDIFLYLPQPLRHKHKSQIDYIPDFLEOLFHEINDLH 373  
319 -----TRVRGGDILYFDMTINNK-----YNL-----FMKNETMYADNH 352  
374 KDLIPVWIAFKTGLRISDVL-----TLQNNCLAKYNGKYSIIITDIATKVGHRIPDN- 428  
353 STEDIYAIGLRE-----QTKDIN--DNIIFOI--OPMNTYYIASQIFKSNF-----GENI-- 400  
429 KLADIITAVLIADSKSKTKDNNPNNYIFAIYKGRKGMFTQHMVRAHLNHLSTKNIID 488  
401 --SGICISIGYRFLGGDWYRHNYLPTVKQG-----NYASLLEST 439  
489 EQGEIFPKTHQF-----RHTYAVKLLNGGADILITQELLAHSSPEWTLAYAKLLDDT 541

RESULT 5  
US-10-092-411A-3242  
; Sequence 3242, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 3242  
; LENGTH: 1151  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-3242  
Query Match 4.9%; Score 116.5; DB 6; Length 1151;  
Best Local Similarity 17.9%; Pred. No. 0.81;  
Matches 97; Conservative 90; Mismatches 175; Indels 181; Gaps 27;  
Qy 3 IPNFISYTNNSLLKDIINEYFNINDSKILSNRKNLTIV---DTSGYNAEYSEGDV 58  
Db 72 LPSSVPPTISNTAIVSAKE-----DEILANTQQQYTKLTIYNDVKOMRQNVDSI 125  
Qy 59 QUNPIPPFD--FKLGSGEDRGKVI-----IVTON---ENIYVNSWYSEFSIS 100  
Db 126 EIN--FPSKNFNLGTFQDTFGVDSNVATHOSTEQPMYTSNLLFYTVYDNYHVSALNFK 183  
Qy 101 FWIRINKWVSNLPGYTIIDSV-----KNNSGWSIGI----- 131  
Db 184 FMKRLNE--VGPIIFVINQDKHNEBELITETFKSRVEKSIKDWDIKLODTYVVSFKDH 241  
Qy 132 -----ISNLFVFTLKQNEDESEQINSFYDISNAPGYNKWFVTVTNMGMNKIYIN 184  
Db 242 PQNEIDKLSNLFVMDQHRESTEDYVNETIQFITDAQ-----YIYIONEMQ----- 287  
Qy 185 GLIITIKVKELTGINFESKT--ITFEINKIPDTGLITSDSDNNMWRDIFYAKELDKGD 243  
Db 288 -SLDPTLQINE--RQFEAYIQOQNO-----EVSAAQQLN-----D 322  
Qy 244 INILFNSLOYTVNWKYDWGN-----DLRYNKEYYMWNI--DY--LNRYMYANSQIVTENT 294  
Db 323 SNOLFNYLKQKR--KDILDNAYIMTYDMRESILRYLESMTATDFKVNCF-----FNK 371  
Qy 295 RNNNDNEGYKIIIKIRGTNDTRVRGG-----DILYEDMTINNKAYNLFMKNE 345  
Db 372 RKKKEEQ-----IKRLNEATQLOEKYVNOVROPLREDSFTRFNKAVNEKILNQ 425  
Qy 346 -----TMYADNHSTEDIYAIGLREQT--KOINDNIIFQIOPMNTYYIASQIFKSN 394  
Db 426 EYDVVPSLISELYQTSISNTYVLTFSDEVIKALNKKIENESTP-----LFEEA 475  
Qy 395 FNGENISGICS-----IGYRFLGGDWYRHNYLPTVKQGNYSALLSTS 440  
Db 476 VNHQVYNELSSDENEDRYEYRIELNTLKDSLTSHTNKHYYI---HLDDSLDKLIGTE 532  
Qy 441 THW 443  
Db 533 THF 535

RESULT 6  
PCT-US02-05068-98  
; Sequence 98, Application PC/TUS0205068  
; GENERAL INFORMATION:  
; APPLICANT: Diversa Corporation  
; TITLE OF INVENTION: ENZYMS HAVING ALPHA AMYLASE ACTIVITY  
; FILE REFERENCE: 09010-108W01  
; CURRENT APPLICATION NUMBER: PCT/US02/05068  
; CURRENT FILING DATE: 2002-02-21  
; PRIOR APPLICATION NUMBER: US 60/270,495  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: US 60/270,496  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: US 60/291,122  
; PRIOR FILING DATE: 2001-05-14  
; NUMBER OF SEQ ID NOS: 321  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 98  
; LENGTH: 976  
; TYPE: PRT



ORGANISM: Bacterial  
PCT-US02-05068-98

Query Match 4.5%; Score 108; DB 1; Length 976;  
Best Local Similarity 21.8%; Pred. No. 2.4;  
Matches 101; Conservative 70; Mismatches 157; Indels 136; Gaps 27;

QY 2 TIPNI-FSTNNSLLKDI--NEVENINDSKILSLQNRKNTLVDTSGYNAEYSEGD 57  
DB 206 TVKEVDLKDASMLPEKVLHNEFIEVLPKAKSL--TENVVDT-----EVNEP-- 256  
QY 58 VQLNPFP-----FDFKLGSGEDRGKIVIVTQNEINIVNSYFISFWI-----RIN 106  
DB 257 -KLWVFPQIELFPDPTNLGSE-----INELEGNLSMNWDYSDPNABFYVYKLEBQ 310  
QY 107 KWSNLPD-----YTIIDSVKN--SGWSGIISN---FLVETLQN-----EDSE 147  
DB 311 EYLFEVGKTRKSYTIENTKOEFPKFSIAINVTANKESGLVVLKKNKLLDLESV 370  
QY 148 QSINFSYDISNAPGVNKFV--TVTNMMGMNKIYINGKLIDTIKVKELRGINFRTI 205  
DB 371 DSISATYVDTNELKLD-WNYTSSVTFEV---LKKGINSVEYII--SOLTQNSFSTEF 424  
QY 206 T----FEINKIPDTGLTSDSDINMWIRDFYFAKELDGKINI-----LFNSIQYTN 255  
DB 425 TGRFVLEKIAIRVANGFESKIN-----EISRDDITISLNLPLTSSMTMYL 473  
QY 256 VKDYWGD-----KNETMYADNHSSTEDIYAIGLEQT-KDIN-DN 372  
DB 474 FRSYFDTDGDGVGDFSGVAEKVDYLSGLVDVWVLPENPKSKSYHGVDYDVEPDY 533  
QY 279 LNRVYANSQIVENTRRNNDNEGYKIIIRKIRGNTDTRVGGDILYFDMTINNAY 338  
DB 534 GTLODLNMIAKVL-----NEN-----GIKVYMDLVVNHSTDPWFELDAY--ENTTNSPY 582  
QY 339 NLFM-----KNETMYADNHSSTEDIYAIGLEQT-KDIN-DN 372  
DB 583 NYIIMSLDEPQKNHWHYKNSGQIVWYVGLFDSMPLNDN 626

RESULT 7  
US-09-721-456-189

GENERAL INFORMATION:  
APPLICANT: Cech, Thomas R.

Lingner, Joachim  
Nakamura, Toru  
Chapman, Karen B.  
Morin, Gregg B.  
Harley, Calvin B.  
Andrews, William H.

TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
NUMBER OF SEQUENCES: 727  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
City: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/721,456  
FILING DATE: 22-Nov-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,549A  
FILING DATE: 19-NOV-1997  
APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: US 08/911,312  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: US 08/912,951  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: US 08/915,503  
FILING DATE: 14-AUG-1997  
APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-OCT-1997  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph Ted  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002610US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 189:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 897 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 189:  
US-09-721-456-189

Query Match 4.5%; Score 107.5; DB 5; Length 897;  
Best Local Similarity 18.5%; Pred. No. 2.3;  
Matches 101; Conservative 80; Mismatches 174; Indels 191; Gaps 26;  
QY 7 IFSY-TNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGY----- 48  
DB 217 LFSYTDNKCVTQFINFEFFYNILPKDLTGRNKNFKQKVKYVELNKEHLHKLLLEK 276  
QY 49 --NAEVS-BEGDVQLNPFPDFK-----LGSSGDEGRGVIVTQNEINIVY-NSMYES 96  
DB 277 INREISWQVETSAAKHFYFDHENIYVLKLLRWTFD---LVVSLRCFFYVTEQKS 333  
QY 97 FSIWFIRINKW-----VSNLPGYTIIDSVKNNSGWSIGIISNPLVFTLKQNEDESQS 149  
DB 334 YSKTYYYRKNIDVIMKMSIADLKETLAEVQE-----KEVEENKS 375  
QY 150 INFSYDISNAPG-----YKNWFFVTVTNMMGMNKIYINGKLIDT-IK 192  
DB 376 LGP-----APKRLIPKKTTFRPFIMFNK-----KIVNSDKTKTLTNTKLLNSHLM 424  
QY 193 VKELTGINFSTITFEINKIPD-----TGLITSDSNDIN--- 226  
DB 425 LKTLKRNMFDPFGFAVENYDDVMKKYEEFVCKWQVGPQLFFATMDIEKCYDSVNRK 484  
QY 227 -----MWIRDFYFAKELDGKIDINILFNSLOY-TNVVXDYWGNDLRYNKYYWNI 276  
DB 485 LSTFLKTKLLSSDFWIMTAQILKRNKNIIVDSKPKKEMKDY-----FKQFKQIAL 538  
QY 277 DYLNRYMANSRQIVENTRRN-NNDFNEGYKIIIRKIRGNTDTRVGGDILYF----- 329  
DB 539 E-----GGGYPTLFSVLENEQNDLNAKTLIVEAKQNR-----YFKKNDLL 579  
QY 330 DMVINNKAYNLFKNKNTWYADNHSTEDIYAIGLEQTKIDINNIIFQTPMNTYYASQ 389  
DB 580 QPVINICQYNYINFNGKFY-----KOTKGIPOGLC--VSSILUSSFYIATL 622  
QY 390 -----IFKSNFNGENISGICSGTYRFRFLGDDWYRHNLYVPTVRKQGNYSALLF---ST 439



SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/721,456

FILING DATE: 22-NOV-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/974,549A

FILING DATE: 19-NOV-1997

APPLICATION NUMBER: US/08/724,643

FILING DATE: 01-OCT-1996

APPLICATION NUMBER: US/08/844,419

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US/08/846,017

FILING DATE: 25-APR-1997

APPLICATION NUMBER: US/08/851,843

FILING DATE: 06-MAY-1997

APPLICATION NUMBER: US/08/854,050

FILING DATE: 09-MAY-1997

APPLICATION NUMBER: US/08/911,312

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US/08/912,951

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: US/08/915,503

FILING DATE: 14-AUG-1997

APPLICATION NUMBER: WO PCT/US97/17618

FILING DATE: 01-OCT-1997

APPLICATION NUMBER: WO PCT/US97/17885

FILING DATE: 01-OCT-1997

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph Ted

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-0026100S

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 110:

SEQUENCE CHARACTERISTICS:

LENGTH: 1031 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: Protein

SEQUENCE DESCRIPTION: SEQ ID NO: 110:

US-09-721-456-110

Query Match 4.5%; Score 107.5; DB 5; Length 1031;  
Best Local Similarity 18.5%; Pred. No. 2.8;  
Matches 101; Conservative 80; Mismatches 174; Indels 191; Gaps 26;

QY 7 IPSTY-TNNSLLKDIINEYFNNI-----LGSSGSDRGKVIIVTQNEIIVY-NSMYES 96  
DB 348 LFSYTTDNKCVTQFINPEFYNLPKDFLTGRNKNFQKKYKVELNKHHLHKNLLEK 407  
QY 49 --NAEVS-EGDVQLNPFPDFK-----LGSSGSDRGKVIIVTQNEIIVY-NSMYES 96  
DB 408 INTRISWQVETSAKHYYFDHENIYVLKLLRWIFED---LVVSLIRCFYFVTEQOKS 464  
QY 97 FSIWIRINKW-----VSNLPGYTIISVKNNSGWSIGIISNLFVFLTKONESEQS 149  
DB 465 YSKTYVYRKNIWDVIMKMSIADLAKETLAEVQE-----KEVEEWKKS 506  
QY 150 INFSDYISNAPG-----YNKWFVTVTNMGMNKIYINGKLIDT-IK 192  
DB 507 LGF-----APGKURLLPKKTTPFPIMTFNK-----KIVNSDRKTYKLTNTKLLNSHLM 555  
QY 193 VKELTGINFSTIITEINKIPD-----TGLITSDSDNIN--- 226  
DB 556 LKTLNRNFKDPFGFAVFNVDYDKVKEEYFVCKWKVQGPQLFATMDIEKCYDSVNRK 615  
QY 227 -----MWIRDIYIFAKELDKDINILNSLOY-TNVVKDYWGNDLYNKEYYMWNI 276  
DB 616 LSTFLKTLKSSDFWIMTAQILKRNKNIIVDSKNFRKKEMKDY-----FRQFKIAL 669  
QY 277 DYLNRYMYANSQIVFNTRRN-NDNFNEGKIIRKIRGNTNDFVRGGDILYF----- 329

Db 670 E-----GGQYPTFLSVLEQNQDLNNAKKTILIVEAKORN-----YFKKDLML 710  
QY 330 DMTINKAYNLKFMKNETWYADNHSTEDIYAIGLREQTKDINDNIIFQIOPMNTYVYASQ 389  
DB 711 QPVINICQYNYINFNGKFY-----KOTKGIPOGLC--VSSILSSFYATL 753  
QY 390 -----IFKSNFNGENISGICIGTYRPLGGDWYRHHYLVPTVKQGNVYASLLE---ST 439  
DB 754 EESSLGLPLRDESMNPENPN-----VNLMLRLTDD-----YLLITTCENNAVLFIKLI 803  
QY 440 STHWGF 445  
DB 804 SRENGF 809

RESULT 10

US-10-092-411A-5178

Sequence 5178, Application US/10092411A

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

FILE REFERENCE: 032796-101

CURRENT APPLICATION NUMBER: US/10/092,411A

PRIOR FILING DATE: 2002-03-07

PRIOR APPLICATION NUMBER: US 09/134,001

PRIOR FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5676

SEQ ID NO 5178

LENGTH: 1010

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-10-092-411A-5178

Query Match 4.4%; Score 106; DB 6; Length 1010;

Best Local Similarity 19.7%; Pred. No. 3.4;

Matches 75; Conservative 60; Mismatches 100; Indels 146; Gaps 19;

QY 8 PSYNNLSLLKDIINEYFNNI-----NDSKILSLQNRKNTLVDTSGYNAEYSEGDVQLNPI 63  
DB 674 YEYENK-TKDKLNLNKKLKIEMDQKHLT-ENLTQTSKEIN--NLELMEKEMQ---- 725  
QY 64 PFDFKLGSSGSDRGKY---IVTQNEI-----VYNSMYESFSISFWIRINKWVSNLPGY 115  
DB 726 -----QLGFESYDQVKSADLSAQKDELEIREINLYNKYQSYE---IEINRLKELVKCK 776  
QY 116 TIID-----SVKNNSGWSIGIISNLFVFLTKONESEQSINFSYDISNAPCYNKWFV 169  
DB 777 KLLNLEELRQSIKTN-----LKLDETNSQIATISYKIDNNSNKNFK---- 818  
QY 170 TVTNMGMNKIYINGKLIDTPIKVKELTGINFSTIITEINKIPDTGLITSDSDNINMWI 229  
DB 819 -IKN-----LIQIDDELKQV----- 833  
QY 230 RDFYIFAKELDKDINILNSLOYTNVVKDYWGNDLYNKEYYMWNIIDLYNRYMYANSQ 289  
DB 834 KEIFLSSLEILAGK-----NDYKLTLENYVL-IYLEKLIQFQANOR 872  
QY 290 IVFNTRRNNDNFNGYKTIIRIRGNTNDFVRGGDILYFDIMINNKAYNLFMKNETMYA 349  
DB 873 LSFMS-----GNRYQL-----IRRETISLGLGLEIDVFD-----FHSNKRHI 911  
QY 350 DNHSPTEDIY-----AIGLRE 364  
DB 912 SLSLGGEFQASLALGLSE 932

RESULT 11

```
US-10-092-411a-3159
; Sequence 3159, Application US/10092411a
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092.411a
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411a-3159

Query Match      4.4%; Score 105; DB 6; Length 10182;
Best Local Similarity 21.2%; Pred. No. 71;
Matches 121; Conservative 71; Mismatches 202; Indels 176; Gaps 28;

QY 6 NPSYTNNSLLDIINEYFNNDKILSLONRKN--TLVDTSGYNAEVSEEGDVQLNP- 62
D 1313 NVSSVTVTRKGDILKEFSSNNNNFLRWNIIRDYIISTSSKIQGRNDDGGLNSPD 1372
QY 63 -----IPFPFKLGSGGEDRGKVIYVQ-----NENIVYN-----SMYSEFSISFWIRI--- 105
D 1373 HKSLIYKDYTLGRQINTNDVLTLLQATAKNSLRSNINSNEKQALAEGRSGNGYSKSIIRD 1432
QY 106 ---NKVSNLPGYTIIDSVKNNSGWSIGIISNLFVTLKQEDSEQISNFSYDLSNAPG 162
D 1433 DGEKSYLLNSPIQVLDLVEPDNGGGRQVSHSNVI---YNEKSSIVNGQVPPANGASA 1489
QY 163 YN--KWFVTVTN--MMG-----NMKIYINGKLIDIRY 193
D 1490 ENIDKVVKANAANGIMGVIYKAQLYAPSPKGYIEKLQNLSTNNVINVYFVPSDKV 1549
QY 194 KE-----LTGINFSTIIFENKIPDPTGLITSDSDNINNMWTRDIYIFAKELDG 241
D 1550 NPSITVGNVDHHTVYSGETFKNTINVNDYGLNVTASTSDS-AITMTN-----NELVG 1603
QY 242 KDNILFNSLQVTNVYK--DYWGNDLRYNKEYVMNIDYLN---RYMANSRQI---VFN 293
D 1604 QAPNVI-NSINKTVKATDKSGNESIYS---FTYNIKPLNEKYRIITSSNQTPVRISN 1659
QY 294 TRNNN-----DEN-----EGYKIIKIRGNTNTRVGGDILYFDMTINKAYNLFMK 343
D 1660 IQNANLSIEDQNRVKSSLSMTKILGTRNVYNESNDVRQSVYKVNRSNGNATVNV--- 1716
QY 344 NEMWAD-----NHSDEDIYA----- 359
D 1717 -TTTFSDGTNTITVPVKHVLLEVPTTTRTVGGQFTFGKTSNDFSLRTGCPVDAR 1775
QY 360 -IGLREQTKDINDN-----IIF--QIQPM-NNTYYAS-----QIFKSNFNGEN 399
D 1776 IYVWNNQGPDIINSQIGRDLTLHAEIFFGGETTPRKDITVYKLSQISIPKQIYETTING-- 1833
QY 400 ISGICSGIYRFLGWDGRHNYLVPTVKQ 429
D 1834 -----RFNSSGDPAIGNF-VQAVNQ 1852

RESULT 12
US-10-092-411a-5114
; Sequence 5114, Application US/10092411a
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
```

```
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092.411a
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 5114
; LENGTH: 497
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411a-5114

Query Match      4.4%; Score 104.5; DB 6; Length 497;
Best Local Similarity 21.3%; Pred. No. 1.7;
Matches 83; Conservative 51; Mismatches 153; Indels 103; Gaps 20;

QY 22 EYFNINDSKILSLONRKNLTVDTSYNAEVSEEGDVQLNPFFDFKLGSGGEDRGKVI 81
D 129 EGFNLYHFNDLTQLDKH-----DVSGYK-----FNLESIASINFTSGTGPQKAVPQ 175
QY 82 VQQENIVYNSMYSEFSISFWIRINKWVSNLPGYTIIDSVKNNSGWSI---GIISNLFV 138
D 176 TFNNHLASAKGCKQSGLGFE---QNTVLSVLPYIHI-----SGLSVILRAVIEGTVR 225
QY 139 TLK--QNEDESGISNFSYDISNNA--PGYNKWFVTVTNMMGNMKYING-----KLID 189
D 226 LVKKTQTDMLTQIK-TYPITHMSLVPTQLKWLMDAGLTQPSLEKILLGKAKLSPOLIE 284
QY 190 ---TTKVKELTGINFSKTITFEINKIP-----DTGLITSDSDNINNMWTRDIYIFAK-E 238
D 285 QALTVELPYNSFGWTEFCQSFLTASQOMLKERFDT--VGKPSENVKIKNPAYGHGE 342
QY 239 LQGDINILFNSLQVTNVYKYGWNDLRYNKEYVMNIDYLNRYMYANSRQIVENTRNN 298
D 343 LLIGENVM-NGLYPKYLKDTFND-GYFQTGDAEID----- 379
QY 299 NDFNEGYKIIKIRGNTNTRVGGDILYFDMTINKAYNLFMKNETMYADNHSTEDIY 358
D 380 ---DEGYVIYDR---RKDLISGGENIY-----PYQI-----ETIAKDFEGIEDAV 420
QY 359 AIGLRBQT-----KDINDNIIFQ 376
D 421 CVGISDDTWGQVPILYVYVNTQDINQTELIE 450

RESULT 13
US-10-092-411a-4341
; Sequence 4341, Application US/10092411a
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092.411a
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 4341
; LENGTH: 695
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-10-092-411a-4341
```

Query Match 4.4%; Score 104.5; DB 6; Length 695;  
Best Local Similarity 21.8%; Pred. No. 2.7; 138; Indels 161; Gaps 25;  
Matches 100; Conservative 60; Mismatches 161; Gaps 25;

QY 10 YTNLLKDIINEY-----FNNINDSKIL-SLQNRKNTLVDTSGYNAFVSSEGDVQLN 61  
DB 185 YGDDNVLKDIAPYEAIRFLTKDKIKDKPLNLYKSHFDQRKYRQLES-----237  
QY 62 PIFPDPFKLSSGDRGVIVTQENIVYNSMYESFISFWIRNKWSNLPQYTIID-- 119  
DB 238 -----TFKCIYNGE-----FLNNVTFSSQHVKKIHRHLGSEKOIDIA 279  
QY 120 SYKNNSGWSIGIISNLFV-----TLKQNEDESESI-----NFSY 154  
DB 280 SRLKESRWMLIRKNKYLNFCKTYEKRTHMSYIKNLQIPVDDDSIFTESPHGNFSG 339  
QY 155 DISNAPGY-----NKKFFVTVNNNM-----GNMK-----180  
DB 340 D-----PKYIALAKRQVDHKIIYVSTSLVDMEIKRYGFTPVRFSGSEKYIKTFKCKY 394  
QY 181 IYING-----KLDDTIKVKELAGINFSTIITPEINK-----IPDTG-----LI 218  
DB 395 VFIENGSDKVVYKSSDQIFVQTHGFPPLKMWV-DLNEQHERQQOLEAFIPRMKKWDYIL 453  
QY 219 TSDSNINMWIRDFVIFAKELDGDKNILFNSLOYTNVVDYV--GNDLRYNKEY--YM 273  
DB 454 TSDSINTTILESAPML-----NKNPNL--KVLEYGAPKNEYLINNNNLOERQQLQKYM 505  
QY 274 VNIDLRYMY-----ANSRQIVFNTRRNND-----FNSGYKIIRK-----RIRGN 315  
DB 506 YKIDDDKKVILYCPWRGNQKEV--TQINKLKKLLKYLPEYELIVKLHPNESHLTRYN 563  
QY 316 TNDTRVGGDILYFDMTINNKAYNLFKKNETMYADNHST 354  
DB 564 QIDNRH-----CYFNELVD--IQELYILSECMTIDYSST 596

RESULT 14  
US-10-092-411a-5080  
; Sequence 5080, Application US/10092411a  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411a  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 5080  
; LENGTH: 3696  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411a-5080

Query Match 4.3%; Score 102; DB 6; Length 3696;  
Best Local Similarity 19.4%; Pred. No. 31;  
Matches 92; Conservative 69; Mismatches 166; Indels 148; Gaps 21;

QY 6 NIFSVYNN---SLKDI--NEYFNINDSKILSLQNRKNTLVDTSGYNAEVS 53  
DB 238 NGFTWYNGKVLGSLNSVLETSVFSADPKYQADINNVVALGRKGNPDNDHGFN---G 294  
QY 54 EESDVOLNP-----IPFDFKLGSSGEDRGKVIYVTONEN-----IYNSMYESFISFI-- 99  
DB 295 IREKFSVNPSEIIFSFNTMTAKRKGQTQLVLRNAENQBIASDIQGGGVYRFLKPLD 354  
QY 100 -----SFWIRNKWSNLPQYTIIDSYKNNSGWSIGIISNLFVLTKQN 143

---

DB 355 NVRLKVCQFLPMEIHSDFKRIQQLHDGYYYSFIDTIGVNSGSHLYVKSQVKNKVG 414  
QY 144 EDSQSI-----NFSYDISNAPGYNKWEFFVTVTNNMGNMKIYIYNGKLIDTIKVKELT 197  
DB 415 KEFEVNTRIENNGFAAAGQNELTYK-----VTLPENFE-----YVD-----NSTEVSFYN 461  
QY 198 GINFSTKIT-PEINKIPDTGLITSDSNINMWIRDFVIFAKELDGDKNILFNSLOYTNV 256  
DB 462 GNVFNSVTFVSVNFDRQNHDTLTFSSNGLNG-RSADQVARELPNKLNI-----510  
QY 257 VKDWGNDLRYNKYYMVNIDYLNRYMYANSROIVFNTRRNNDNFEGYKIIIKRTRGWT 316  
DB 511 -----RY--KLRFVNI-----STPREVTFNEAIKYTFSEYI-----NT 543  
QY 317 NDRVTRVGGDILYFDMTINNK-----AYNLFMK-----N 344  
DB 544 NDNTVTGQOTPFSSINVINKDDISEQVKNKDIIPSNYTLASYNKYKLERACTVLDEEN 603  
QY 345 ETMYADNHSTEDIVAGLRBOT-----KIDNMIIFQIQPMNTYTYASQI 390  
DB 604 NTPFNQRYSQIIDDLLHELQTTLINVSASREIND-----KAOEMTDAVYDSTEL 654

RESULT 15  
US-10-092-411a-4465  
; Sequence 4465, Application US/10092411a  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411a  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 4465  
; LENGTH: 527  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411a-4465

Query Match 4.2%; Score 100; DB 6; Length 527;  
Best Local Similarity 20.1%; Pred. No. 3.7;  
Matches 93; Conservative 73; Mismatches 187; Indels 110; Gaps 22;

QY 4 PNIFSVYNNLLKDI---INEYFNINDSKILSLQNRKNTLVDTSGYNAEV---SEEGD 57  
DB 117 PFQIQATIGDNTFFSKIHFSQGYLMWVEDYKYSTIQRR--FVDDRGRFISAVRTYDGD 174  
QY 58 VOLNPIFPDFPKLGSS--GEDRGKVIYVTONEN-----IYNSMYESFISFWIRNKW 108  
DB 175 NKKHYFSKD---GEEIFVEDLVNTVTINKNFQSKFKRVTYSSMAELIEKFQSYVERE 231  
QY 109 VSNLGYIIDSVKNNKSGWSIGIISNLFVTL--KQNEDESEQISNFSYDISNAPGYNKW 166  
DB 232 LNEEDSVIVASDERHNSMMATIDASSLCFSIFTERKNKVTQDL---YDSISRA-----282  
QY 167 FFVTVTNNMGNMKIYIYNGKLIDTIKVKELTGINFSTIITPEINKIPDTGLITSDSNIN 226  
DB 283 YYCLVDQTQANQM-----IEHYAGLNMNDINLLAVTPFDKSLPQSSQLYDT-YIG 333  
QY 227 MWIRFYIFAKELDGD-----INILFNSLOYTNVVKDWGNDLRYN-KEYYMWYDI 278  
DB 334 LMI-----DGLDEIETREIVNSLFQYIQHKGQYKILTKSRDNLNIDIEVAH 383  
QY 279 LNRYMYANSRQT-----VFNTRRNND-----FNEGYKIIIKRIRGNTNDRVGGDI 326

Db 384 LNDLYHQEKKEISDVIEDVIONKKETIIDETVPFEEDLYSVISKLR----- 430  
QY 327 LYFDMTINNKAYNLFM-----KNETWVADNHSTEDIYAIGLREOTKDINDNI 373  
Db 431 VVVDLSLEPK---LFLQICCGAGIPQINKKRTDYVKH--MHNGYIIDDISQTVESLDYF 485  
QY 374 IFQIQPMNNTYYASQIFKSNGENI-----SGICSIGT 408  
Db 486 LAHLKNWNSYAYSML-TDDFSSINIHQINQLFKGDVSSGT 527

Search completed: November 7, 2002, 14:53:14  
Job time : 20 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 16:32:06 ; Search time 2721 Seconds  
(without alignments)  
4813.033 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIPNFTSYNNSLKDI.....NYASLLESTHNGFVPVSE 450

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delcp 6.0 , Delcxt 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/uspto.spool/US09910186/runat\_04112002\_111617\_20836/app\_query.fasta\_1.647  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09910186@cgn\_1\_1\_1616\_brunat\_04112002\_111617\_20836 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:  
1: gb\_da.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_yi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rnd.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2386	99.8	4479	7	CBCPHGCL X53751 Clostridium
2	2386	99.8	4592	7	CSTCITOX D90210 Bacterioph
3	2386	99.8	4712	1	CBPHCNIM X71126 C.botulinu
4	2386	99.8	9613	1	CBCITOX X62389 Botulinum b
5	2386	99.8	9689	1	CBP1CMTCL X66433 Clostridium
6	2386	99.8	11747	1	AB061780 Clostridi
7	2386	99.8	12297	7	CSCBONT X72793 Clostridium
8	1824	76.3	11631	1	AB037920 Clostridi
9	1823	76.2	4049	1	CLOINCO D38442 Clostridium
10	882.5	36.9	11614	1	CLOBN D49440 Clostridium
11	882.5	36.9	11614	1	AB037166 Clostridi
12	867.5	36.3	3952	1	CBNTTD X54254 Clostridium
13	867.5	36.3	4087	7	S49407 S49407 type D neur
14	867.5	36.3	11584	7	AB012112 Clostridi
15	610	25.5	3876	1	AF300465 Clostridi
16	610	25.5	3876	1	AF300466 Clostridi
17	604.5	25.3	3937	1	CBONTG X74162 C.botulinu
18	604	25.3	11170	1	CBY13630 Clostridium
19	602.5	25.2	3759	1	AB088207 Clostridi
20	602.5	25.2	3838	1	AB039264 Clostridi
21	602	25.2	3876	1	AF295926 Clostridi
22	594.5	24.9	3759	1	AB082519 Clostridi
23	594.5	24.9	4030	1	CBNTTB X82683 C.botulinu
24	594	24.8	3835	1	AB037704 Clostridi
25	594	24.8	3835	1	AB037705 Clostridi
26	594	24.8	3835	1	AB037706 Clostridi
27	594	24.8	3835	1	AB037707 Clostridi
28	594	24.8	3835	1	AB037708 Clostridi
29	594	24.8	3835	1	AB037709 Clostridi
30	594	24.8	3835	1	AB037710 Clostridi
31	594	24.8	3835	1	AB037711 Clostridi
32	594	24.8	3835	1	AB037712 Clostridi
33	594	24.8	3835	1	AB037713 Clostridi
34	594	24.8	3835	1	AB037714 Clostridi
35	594	24.8	3949	1	CBNTTB X62088 C.butyricu
36	593	24.8	4051	1	CBONTB X71343 C.botulinu
37	586.5	24.5	3891	6	AR000031 Sequence
38	586.5	24.5	3891	6	AR169142 Sequence
39	586.5	24.5	3891	6	AX036248 Sequence
40	586.5	24.5	4292	1	CBOTAG X52066 Clostridium
41	586.5	24.5	4835	1	CLONEUR M30196 C.botulinu
42	586.5	24.5	15039	1	AF461540 Clostridi
43	583	24.4	4017	1	CNEUTTOXE X62089 C.botulinu
44	583	24.4	4017	6	AX088262 Sequence
45	581	24.3	4041	1	ME1186 Clostridium

ALIGNMENTS

RESULT 1

CBCPHGC1	CBCPHGC1	DB:	7	Gaps:	0
LOCUS	Clostridium botulinum C phage gene for C1 neurotoxin.	US-09-910-186A-10 (1-450) x CBCPHGC1 (1-4479)			
DEFINITION	4479 bp DNA linear PHG 12-SEP-1993				
ACCESSION	X53751				
VERSION	X53751.1 GI:14905				
KEYWORDS	C1 neurotoxin; neurotoxin; toxin.				
SOURCE	Clostridium botulinum C phage.				
ORGANISM	Clostridium botulinum C phage				
REFERENCE	1 (bases 1 to 4479)				
AUTHORS	Popoff M.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (03-JUL-1990) Popoff M.R., Institut Pasteur, Anaerobies,				
REMARK	25 rue du Docteur Roux, 75724 Paris Cedex 15, France				
AUTHORS	revised by [3]				
TITLE	2 (bases 1 to 4479)				
JOURNAL	Hauser D., Eklund M.W., Kurazono H., Binz T., Niemann H.,				
MEDLINE	Gill D.M., Boquet P. and Popoff M.R.				
PUBMED	Nucleotide sequence of Clostridium botulinum C1 neurotoxin				
REMARK	90370487				
REFERENCE	revised by [3]				
AUTHORS	1. 4479				
TITLE	Location/Qualifiers				
JOURNAL	1. 4479				
FEATURES	source				
source	/organism="Clostridium botulinum C phage"				
source	/strain="phage from 468C botulinum C. strain"				
source	/db_xref="taxon:12448"				
source	198..203				
source	/note="pot. ribosome binding site"				
source	214..4089				
source	/note="C1 neurotoxin (AA 1-1291)"				
source	/codon_start=1				
source	/transl_table=1				
source	/protein_id="CAA37780.1"				
source	/db_xref="GI:14906"				
source	/translation="MPITINNFNSDPVDNKNILYLDTHLNLANPEKAFRITGNIM				
source	VIPDRFSNSPNLKNPRTVSPKSGYDNPSTLSDKDPFLKEIKLIFKRINSREI				
source	GEELYRLSTDPFPPNNPTINTDFDVPNSVDVKTROGNWVKTGINSIPSVITG				
source	PRENIDPETSTKLTNTAAQFGALSTISPREMLTYSNANDVGEGRFSKE				
source	KCMDDILLHELKHMENLYGIAIPNDQISSVTSNIFSOYNKLEAIEYAEQGP				
source	TIDLIPKARFYFEKALDYRSIAKRLNITSNPSPFNKIYGEIKQKLIRKRVV				
source	ESSGEVTVNRKVELYNELTQIEFENYAKIYNQNRKIYLSNVPTVNTLDDNV				
source	YDIQGNFNPISNLNLFMGQNLSPALRKVPNNMLYLFKECHKAIDGRSLNKT				
source	LDICRELLVNTDLFPDIFGDSVKTDFLRKIDNEETEVYIPDNVSDVQVILSKNTE				
source	HGOLLVPSIDSESEILPGENVQVYDNQNVYDLSNYYLESQKLSNVDEFTFR				
source	STEEALDMSAKVTYFPTLANKVAGQGLFLMWANDVDEFTNLRKTDKTSQ				
source	VSALTPTGPAIINSVSRGNFTEAFVGVTVLLEAFPEFTIPALGAFVIYSKQOE				
source	NRKILKIDNLEORIKRWKDSYEMWGMGLSRITITOFNISIQMDSLNQAGAKA				
source	KIDLEBYKSGSDENKSNQVKNLSDVKISAMNNKIFRECSVYLFKRMILPK				
source	VIDELNEFDRTKAKLINDSHIILVGEVDKLRKAVNNFQNTIPFNLFSTNNSL				
source	LKDIENFENINDSKILSLQNRKNTLVDSGNAEYSEEGVQLAPIFPPFKLGS				
source	GEDKGVTVQNIENISVSEFSISFWIRKWNKSNLPCYTIIDSVKNWSGSI				
source	ISNFTVFLKQNESEQISNFSDIINNAPGYNKWFVTVTNNMGMKIYINGKLI				
source	TKVKELGAFSPSTIPEINLPOTGLIISDSNINMWRDVFIPAKELDGKDIL				
source	FNSLYTVNVYDVGWGNLRYNKEYMNYNIDLYNRYMANSRQIVFNRNNPNFEGY				
source	KIIIRKRGNTNVRGGLDILYFDMTINNKANLNFKNMETMADHNSIEDITFALGR				
source	EQTIDINNIIFIQPMNTYIYASQIFKSNFNGENISGCSIGTFRFRLGGDWYRHN				
source	YLVPTVKQGNVASLLESTSTHWGFVPVSE"				
BASE COUNT	1817 a 455 c 693 g 1514 t.				
ORIGIN					
Alignment Scores:					
Pred. No.:	3,4e-152	Length:	4479		
Score:	2386.00	Matches:	449		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		
Query Match:	99.79%	Indels:	0		







```

Db 3873 TATATAAATGGAATAATTAATGATCTATAAAGTAAAGAACTAACTGGAATTAATTT 3932
QY 202 SerLysThrIleThrPheClnrleasnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3933 ACRAAACTATPAACATTTGAATAATAATAATTCAGATACCGTTTGATCTACAT 3992
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrllePheAlaLysGluLeuAspGly 241
Db 3993 TCTGATAACATCAATATGCGATAGAGATTTTATATATTTGCTAAAGATTAGATGTT 4052
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnIleThrAsnValValLysAspTyrlle 261
Db 4053 AAGATATTAATATATATTAATAAGCTGCAATATCTAATATGTTGTAAGAAGATATTGG 4112
QY 262 GlyAsnAspLeuArgTyrlleAsnLysGlnTyrlleMetValAsnIleAspTyrlleAsnArg 281
Db 4113 GGAATGATTTAAGATATAATAAAGAAATATATATGTTAATATAGATTTTAAATAGA 4172
QY 282 TyrMetTyrlleAsnSerArgGlnIleValPheAsnThrArgAsnAsnAspPhe 301
Db 4173 TATATGATGCGAATCGCAAAATGTTTTTAAATACACGTAGAAAATAATAGCTTC 4232
QY 302 AsnGluGlyTyrlleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
Db 4233 AATGAAGATATAAATATATAAAGAAATCAGAGAAATACAAATGACTAGAGTA 4292
QY 322 ArgGlyGlyAspIleLeuTyrllePheAspMetThrIleAsnAsnLysAlaTyrlleAsnLeuPhe 341
Db 4293 CGAGGAGGAGATATTTATATTTGATGATGCAATATATAACAAGATATAATTTGTTT 4352
QY 342 MetLysAsnGlnThrMetTyrlleAlaAspAsnHisSerThrGluAspIleTyrlleAlaIleGly 361
Db 4353 ATGAAGAATGAACATATGATGCGACATATCATCTACTGAGATATATATGCTATAGGT 4412
QY 362 LeuArgGlnThrLysAspIleAsnAspAsnIlellePheGlnIleGlnProMetAsn 381
Db 4413 TTAAGAGAAACAACAAGATATAATGATATATATATATTTCAAAATACACCAATGAAT 4472
QY 382 AsnThrTyrlleTyrlleAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
Db 4473 AATCTATATATAGCATCTCAATATATTTAAUCAATTTTATGAGAGAAATATTTCT 4532
QY 402 GlyIleCysSerIleGlyThrTyrlleArgPheArgLeuGlyGlyAspTyrlleArgHisAsn 421
Db 4533 GGAATATGTTCAATAGGATCTATCTGTTTAGACTTGGAGGTGATGTTGATAGACAAAT 4592
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrlleAlaSerLeuLeuGluSerThrSerThr 441
Db 4593 TATTGGTCCCTACTGTGAGAGAGAAATTAATGCTTCTATTAGATCAACATCAACT 4652
QY 442 HisTrpGlyPheValProValSerGlu 450
Db 4653 CATTTGGGTTTGTACCTGTAAGTGA 4679

RESULT 4
LOCUS CBCTOX 9613 bp DNA linear BCT 29-JUN-1999
DEFINITION Botulinum bacteriophage genes for HA-17, HA-33, nontoxic components
and C1 toxin.
ACCESSION X62389
VERSION X62389.1 GI:558175
KEYWORDS botulinum toxin; C1 toxin; HA-17 gene; HA-33 gene; Hemagglutinin;
neurotoxin; nontoxic; toxin.
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 9613)
AUTHORS Kimura K., Fujii N., Tsuzuki K., Murakami T., Indoh T.,
Yokosawa N., Takeshi K., Suto B. and Oguma K.
TITLE The complete nucleotide sequence of the gene coding for botulinum
type C1 toxin in the C-ST phage genome

```

```

JOURNAL MEDLINE
PUBMED 222445
REFERENCE 2 (bases 1 to 9613)
AUTHORS Tsuzuki K., Kimura K., Fujii N., Yokosawa N. and Oguma K.
TITLE Nucleotide sequence of the gene for one of the components of
hemagglutinin produced by Clostridium botulinum type C
unpublished
JOURNAL REFERENCE 3 (bases 1 to 9613)
AUTHORS Tsuzuki K.
TITLE Direct Submission
JOURNAL TITLE Submitted (19-DEC-1991) K. Tsuzuki, Sapporo Medical College, Dept
of Microbiology, South 1 West 17, Sapporo 060, JAPAN
COMMENT On Oct 15, 1994 this sequence version replaced gi:40386.
See also X53041 & D90210.
FEATURES
source
1. .9613
/organism="Clostridium botulinum"
/strain="type C Stockholm, c-st"
/db_xref="taxon:1491"
/complement(111..551)
/gene="HA-17"
/complement(111..551)
/gene="HA-17"
/codon_start=1
/transl_table=11
/protein_id="CAA44260.1"
/db_xref="GI:40387"
/db_xref="SWISS-PROT:P46083"
/translation="MSERTFLPFGNYKSLFSDSLYITYSGALSFNTSLDNQK
WKLEYSSNGFRFSNVAEPKYLAYDYGFIYLSSSNNSLWNPKIANSIYIICL
SIVNVDYAWTIYDNNNNITDQPIILNPLNPDINNNOILKLEKL"
/complement(560..565)
/complement(614..1474)
/gene="HA-33"
/complement(614..1474)
/gene="HA-33"
/codon_start=1
/transl_table=11
/protein_id="CAA44261.1"
/db_xref="GI:40388"
/db_xref="SWISS-PROT:P46084"
/translation="MSQTANDLRNNEVFIFSPSNNNTNKVLKISQSEVKLNKLSGA
NQWRLIYDNTKQAYKIKYMDNTSLITLWNPASSVSKDITNGDQNYTLQNTLSR
NVIIRYNNPVLQYNIIDITLMVSTQSSNQFFKFSNCIYEALNNRCKLQTLNS
DRYLSKLNQYIIVLQWQFDSRQKWIETNKSAITKCCQNNRYLTIQNNYIV
ETYSTDSLQYNNIYLDNASKYILYLNQDTRVLDVNSQIANGTHVIVDSYHGN
TNOQWITNLJ"
/complement(1482..1486)
RBS 1726..1731
RBS 1737..5327
gene /gene="nontoxic"
1737..5327
/gene="nontoxic"
/codon_start=1
/transl_table=11
/product="non-toxic components"
/protein_id="CAA44262.1"
/db_xref="GI:40389"
/db_xref="SWISS-PROT:P46081"
/translation="MDINDLNTNSPYDNKNVYIVRARKNTTEFFKAFKVPNIWAPE
RYGPELDIAEYKLDGGIYDSNPLSODSERENFLQAIILLKRINNTISGKQLSLI
STAPFPYGVITGGYSSPNITFGKTPKSNKLSLVSTIPFPFGYRETNYIESQN
NKFYASNIIFGPGSNIVNNVYKNDAGMGMTARIVFQPLLTKYKNFYIDP
AWELTKLISLFIYIGIKESDNLVYPRLELTEDNKQFSQNLIDLLISGVVDLEFI
NTNPWTNYSFPNSIKMFYKNYKYTELEGNNAIGNDKLKLKQFQINQVODIWNL
NLNTFCQSFNSIIPRESNALKHFYKQYTDYTDYNYINGVNGQITKLPISNKN
TNIISKPKVNLVNNENIILKMSNIYDGLKGTEDFYSTYKIPNEYEYFNDS
TNENFISIEVDSPETIDINPKNSDNLTQITSMTEETVTTALSINYLQAI
NFPNLNLTSSDFSVSSKSLVSLFDNLMSYLETKNDGPIDTKYLVKLVKVF
KNYSPDNLQEDISMGINEVLMWPKALNLTNSFVEEYQDSALISLKKNL
REPNIIDDSLSLGLSFKDLNKNKLYIVFKLYFSLQWWTYYSQYFE
LICMAKOSILAQESLVKQIVONKFTDLKASIPDITKLIRETEKTFIDLSNESQIS

```

```

Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)
91024998
222445
2 (bases 1 to 9613)
Tsuzuki K., Kimura K., Fujii N., Yokosawa N. and Oguma K.
Nucleotide sequence of the gene for one of the components of
hemagglutinin produced by Clostridium botulinum type C
unpublished
3 (bases 1 to 9613)
Tsuzuki K.
Direct Submission
Submitted (19-DEC-1991) K. Tsuzuki, Sapporo Medical College, Dept
of Microbiology, South 1 West 17, Sapporo 060, JAPAN
On Oct 15, 1994 this sequence version replaced gi:40386.
See also X53041 & D90210.
Location/Qualifiers
1. .9613
/organism="Clostridium botulinum"
/strain="type C Stockholm, c-st"
/db_xref="taxon:1491"
/complement(111..551)
/gene="HA-17"
/complement(111..551)
/gene="HA-17"
/codon_start=1
/transl_table=11
/protein_id="CAA44260.1"
/db_xref="GI:40387"
/db_xref="SWISS-PROT:P46083"
/translation="MSERTFLPFGNYKSLFSDSLYITYSGALSFNTSLDNQK
WKLEYSSNGFRFSNVAEPKYLAYDYGFIYLSSSNNSLWNPKIANSIYIICL
SIVNVDYAWTIYDNNNNITDQPIILNPLNPDINNNOILKLEKL"
/complement(560..565)
/complement(614..1474)
/gene="HA-33"
/complement(614..1474)
/gene="HA-33"
/codon_start=1
/transl_table=11
/protein_id="CAA44261.1"
/db_xref="GI:40388"
/db_xref="SWISS-PROT:P46084"
/translation="MSQTANDLRNNEVFIFSPSNNNTNKVLKISQSEVKLNKLSGA
NQWRLIYDNTKQAYKIKYMDNTSLITLWNPASSVSKDITNGDQNYTLQNTLSR
NVIIRYNNPVLQYNIIDITLMVSTQSSNQFFKFSNCIYEALNNRCKLQTLNS
DRYLSKLNQYIIVLQWQFDSRQKWIETNKSAITKCCQNNRYLTIQNNYIV
ETYSTDSLQYNNIYLDNASKYILYLNQDTRVLDVNSQIANGTHVIVDSYHGN
TNOQWITNLJ"
/complement(1482..1486)
1726..1731
1737..5327
/gene="nontoxic"
1737..5327
/gene="nontoxic"
/codon_start=1
/transl_table=11
/product="non-toxic components"
/protein_id="CAA44262.1"
/db_xref="GI:40389"
/db_xref="SWISS-PROT:P46081"
/translation="MDINDLNTNSPYDNKNVYIVRARKNTTEFFKAFKVPNIWAPE
RYGPELDIAEYKLDGGIYDSNPLSODSERENFLQAIILLKRINNTISGKQLSLI
STAPFPYGVITGGYSSPNITFGKTPKSNKLSLVSTIPFPFGYRETNYIESQN
NKFYASNIIFGPGSNIVNNVYKNDAGMGMTARIVFQPLLTKYKNFYIDP
AWELTKLISLFIYIGIKESDNLVYPRLELTEDNKQFSQNLIDLLISGVVDLEFI
NTNPWTNYSFPNSIKMFYKNYKYTELEGNNAIGNDKLKLKQFQINQVODIWNL
NLNTFCQSFNSIIPRESNALKHFYKQYTDYTDYNYINGVNGQITKLPISNKN
TNIISKPKVNLVNNENIILKMSNIYDGLKGTEDFYSTYKIPNEYEYFNDS
TNENFISIEVDSPETIDINPKNSDNLTQITSMTEETVTTALSINYLQAI
NFPNLNLTSSDFSVSSKSLVSLFDNLMSYLETKNDGPIDTKYLVKLVKVF
KNYSPDNLQEDISMGINEVLMWPKALNLTNSFVEEYQDSALISLKKNL
REPNIIDDSLSLGLSFKDLNKNKLYIVFKLYFSLQWWTYYSQYFE
LICMAKOSILAQESLVKQIVONKFTDLKASIPDITKLIRETEKTFIDLSNESQIS

```



```

RESULT 5
CBP1CMTCL      9689 bp      DNA      linear      BCT 15-FEB-1999
LOCUS          Clostridium botulinum phage 1C, Chn-14, Chn-33, Chn-138 and BONT/CI
DEFINITION     genes.
ACCESSION      X66433.1 GI:509271
VERSION        haemagglutinin; neurotoxin.
KEYWORDS       Clostridium botulinum.
SOURCE         Clostridium botulinum
ORGANISM       Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
               Clostridium.
REFERENCE      1 (bases 1 to 9689)
AUTHORS       Hauser,D.F., Eklund,M.W. and Popoff,M.R.
JOURNAL       Unpublished
REFERENCE      2 (bases 1 to 9689)
AUTHORS       Hauser,D.F.
JOURNAL       Direct Submission
TITLE         Submitted (25-MAY-1992), D.F. Hauser, Institut Pasteur, Unite des
               Toxines Microbiennes, Institut Pasteur, 28 rue duDocteur-Roux,
               F-75724 Paris Cedex, FRANCE
FEATURES       location/Qualifiers
               1. .9689
                  /organism="Clostridium botulinum"
                  /db_xref="taxon:1491"
                  complement(253..647)
                  /gene="Chn-14"
                  complement(253..633)
                  /gene="Chn-14"
                  /codon_start=1
                  /transl_table=11
                  /product="14kDa protein associated with BoNT
                  /CI-haemagglutinin complex"
                  /protein_id="CAA47057.1"
                  /db_xref="GI:509272"
                  /db_xref="SPREMBL:Q45915"
                  /translation="MSSEFTLPNGNYKIKLSFSDSLYLTYSYGALSFNTSSLDNQK
                  WKLEYTSSNGFFSVAPENPKYLAINDYGYFLSSSSNNLSNLPKIAINSYIICLT
                  SINVTDYANTIVDNNNNITDHLF"
                  complement(642..647)
                  /gene="Chn-14"
                  complement(696..1556)
                  /gene="Chn-33"
                  complement(696..1556)
                  /gene="Chn-33"
                  /codon_start=1
                  /transl_table=11
                  /product="33kDa protein associated with BoNT
                  /CI-haemagglutinin complex"
                  /protein_id="CAA47058.1"
                  /db_xref="GI:509273"
                  /translation="MSQTANDLRNNEVFFISPNNTNKVLDKISQSEVKLWNKLSGA
                  NKWRLLYDNTKQAYIKYMDNTSILTWNAPLSVSVKTDINGDQYWLQNYLISR
                  NVILRYNPNLVLYQNDITLMTSTQSSNQPFKFSNCIYEALNNRCKLQTLNLS
                  DRFLSKNLSSQIIVLQWQFDSRSKWIIEYNETKSAYTLKCOENNRVLTWIQNSNVY
                  ERTQSTDSLLQYWNINYLNDASKYLILNQDTRVLDVYNSQIANGTHVYDVSRYGN
                  TNOQWIINLI"
               gene      complement(1564..1568)
               RBS       /gene="Chn-33"
               gene      complement(1564..1568)
               RBS       /gene="Chn-33"
               gene      join(1803..1808,1819..5409)
               RBS       /gene="Chn-138"
               CDS        1803..1808
               CDS        1819..5409
               CDS        /gene="Chn-138"
               CDS        /codon_start=1
               CDS        /transl_table=11
               CDS        /product="138kDa protein associated with BoNT
               CDS        /CI-haemagglutinin complex"
               CDS        /protein_id="CAA47059.1"
               CDS        /db_xref="GI:4376225"

```

US-09-910-186a-10 (1-450) x CBPICNTC1 (1-9689)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
|||||  
Db 7950 ACAATACCCCTTTAATATATTTTTCATATACTAATAATCTTTATATAAAGATATAATTAAT 8009  
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
|||||  
Db 8010 GAATATTTCAATATATTAATGATTCATAATTTTGAGCCGTACAAAACAGAAAATACT 8069  
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGluValGlnLeuAsn 61  
|||||  
Db 8070 TTATGGGATACATCAGGATATAATGCAGAACTGAGTGAAGAAGCGATGTTTCAGCTTAAT 8129  
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
|||||  
Db 8130 CCAATATTTCCATTGACITTAATATTAGTASTTCACGGGAGGTAGAGGTAAAGTTATA 8189  
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
|||||  
Db 8190 GTAACCCAGAAATGAATATTTGTATATAATCTATGATGAAAGTTTTCAGCTAGTTT 8249  
QY 102 TtpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
|||||  
Db 8250 TCGATTAGATTAATTAATGATGATTAATTTTACCTGGATATACATAATGTAGTAGTT 8309  
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
|||||  
Db 8310 AAAATAACTCAGGTGGAGTATAGGTATATTAGTAATTTTTCAGTATTTTAAAA 8369  
QY 142 GlnAsnGluAspSerGluGluSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161  
|||||  
Db 8370 CAATATGAAGATAGTGAACAAAGATAAATTTTAGTTATGATATATCAATAATGCTCCT 8429  
QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
|||||  
Db 8430 GGATACAATAAATGCTTTTGTAACTGTTACTAACAATATGATGGCAATATGAAGATT 8489  
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
|||||  
Db 8490 TATATAAATGGAAAATTAATAGACTATAAAGTTTAAAGAACTAACTGGAATTAATTT 8549  
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
|||||  
Db 8550 AGCAAACTATAACATTTGAATAATAATAAATTCACAGATACCGGTTTGATCTCAGAT 8609  
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
|||||  
Db 8610 TCTGATAACATCAATATGATGGATAAGAGATTTTATATATTGCTAAAGAATTAGATGTT 8669  
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261  
|||||  
Db 8670 AAGATATTAATATATTTAATAGCTGCAATATACTAATGTTGTAAGAATTATTGG 8729  
QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
|||||  
Db 8730 GGAATATGATTTAAGATATAATAAAGAATTTATATGTTTAAATATAGATTATTTAAATAGA 8789  
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
|||||  
Db 8790 TATATGATGCAACTCAGCAAAATGTTTTTAAACACGTAGAAATATATATGACTTC 8849  
QY 302 AsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
|||||  
Db 8850 AATGAAGGATATAAATAAATAAAGAATCAGAGAAATACAAATGATCTAGAGTA 8909  
QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
|||||  
Db 8910 CCAGAGGAGATATTTATTTTATGATGACAAATTAATACAAAGCATATAATTTGTTT 8969  
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361  
|||||

Db 8970 ATGAAGAATGAACACTATGATGTCAGATAATCATAGTACTGAGATATATATGCTATAGGT 9029  
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
|||||  
Db 9030 TTAAAGAGAACAAACAAAGGATATAAATGATAATATTTATTTCAAAATACAACTGAAT 9089  
QY 382 AsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
|||||  
Db 9090 AATACTTATTATTCAGCATCTCAAAATATTAAATCAAAATTTAATGGAGAAATATTTCT 9149  
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421  
|||||  
Db 9150 GGAATATGTTCAATAGGTACTTATCGTTTAGACTTGGAGTGATGGTATAGACACAAT 9209  
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441  
|||||  
Db 9210 TATTGGTGCCCTACTGTGACCAAGGAATTTGCTTCATATTAGAAATCAACATCAACT 9269  
QY 442 HisTrpGlyPheValProValSerGlu 450  
|||||  
Db 9270 CATTTGGGTTTTGTACCTGTAAGTGAA 9296  
RESULT 6  
LOCUS AB061780 11747 bp DNA linear BCT 06-NOV-2001  
DEFINITION Clostridium botulinum orf-22, ha-70, ha-17, ha-33, ntaha, nt genes  
for ORF-22, HA-70, HA-17, HA-33, NTNHA, neurotoxin, complete cds.  
ACCESSION AB061780  
VERSION AB061780.1 GI:16580753  
KEYWORDS  
SOURCE Clostridium botulinum (strain:C-Yoichi) DNA.  
ORGANISM Clostridium botulinum  
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
Clostridium.  
REFERENCE 1  
AUTHORS Segane,Y., Kouguchi,H., Watanabe,T., Sunagawa,H., Inoue,K.,  
Fujinaga,Y., Oguma,K. and Ohyama,T.  
TITLE Role of C-terminal region of HA-33 component of botulinum toxin in  
hemagglutination  
JOURNAL Biochem. Biophys. Res. Commun. 288 (3), 650-657 (2001)  
MEDLINE 21534265  
REFERENCE 2 (bases 1 to 11747)  
AUTHORS Segane,Y., Watanabe,T., Kouguchi,H., Morita,M., Miyamoto,T.,  
Yanamoto,T., Muroga,A. and Ohyama,T.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-2001) Toshihiro Watanabe, Tokyo University of  
Agriculture, Faculty of Bioindustry, Yassaka 196, Abashiri, Hokkaido  
099-2493, Japan (E-mail:t-watanabe@bioindustry.nodai.ac.jp,  
Tel:81-152-48-3843, Fax:81-152-48-3843)  
FEATURES  
source  
1..11747  
/organism="Clostridium botulinum"  
/strain="C-Yoichi"  
/db\_xref="taxon:1491"  
6..545  
/gene="orf-22"  
6..545  
/gene="orf-22"  
/codon\_start=1  
/transl\_table=11  
/product="ORF-22"  
/protein\_id="BA871744.1"  
/db\_xref="GI:16580754"  
/translation="MNDLFYAIENLKHNDNQNFIMESLUKKYIKTSKKYNLYDYNN  
DIYHLWKELIENLKNFENSELDKRYISTKRYCINICKKNDKKIYNSEVYK  
KLQAVNYSLYCDNFELDLISLTNYKQIIYMKFEGRKDNIEAIRLSRQSIYK  
IRIKSLKKLYPIVMQLVNI"  
complement(620..2491)  
/gene="ha-70"  
complement(620..2491)  
/gene="ha-70"  
/codon\_start=1  
/transl\_table=11

```

/feature="HA-70"
/protein_id="BAB71745.1"
/db_xref="GI:16580755"
/translation="MSISIKELAYTKDKSINNVNLAGVNVVNRGCGWILSRQNLIG
GNISNNGCTAIYGLRLREYATPYPTASFNERYKNNVQVFNFTASPIPGFE
FSKTAPEKSLMYLQTYIIRYELIKVQNTVERAVLVPSGLGVKSLFEPNSEBQID
KNFYTSODKILNEKFIYKIDDTIVTESKSNMNNINFNTSOTILPYNGLYVINK
GDGYRTNDKDLIGTIELIESTGSTSIQPRLRNTTRPLFNTSPTIFSOEYEARLND
AFNTQLENTTLTKFVEEAPTNNKISMKVNTYERKELINPQNGNIDSKAEYILPSL
KCEVSDAPSPQAPVPTVDQDGTQGTQGNENIIVGNTPENSEIEETIPDPDVTY
NIPYSIONNACYLVFKYNTGVYKIKTKNPLPILLYEALGSSNNKNSNINSDNIK
AIKTYIGLNSDAKSYLVLSLEKDKVYIRIPQISSSTTSQIFRKLGNISDLADST
VILNDLNTSGTHYTRQSPDVGNYISYQITIPEDFNFIASSIFSPTKNNOGITLY
RLTISNGYLNITINNYSDDLNNVPEISLLNGATYIFRVKVELNINNIIFDAYRNS"
complement(2505..2945)
/gene="ha-17"
complement(2505..2945)
/gene="ha-17"
/codon_start=1
/transl_table=11
/product="HA-17"
/protein_id="BAB71746.1"
/db_xref="GI:16580756"
/translation="MSSERTFLPNGYKIKLSFNSLYLITYSSGALFSNTSLDNQK
WKLXIYSSNGFRFSVAEPNKYLAYNDGYFVLSSSNNSLWNPRIKAINSYICTL
SIVNVDTYATVYDNNNNITDQPLILNPNEDINNQSQILKLEK"
complement(2999..3856)
/gene="ha-33"
complement(2999..3856)
/gene="ha-33"
/codon_start=1
/transl_table=11
/product="HA-33"
/protein_id="BAB71747.1"
/db_xref="GI:16580757"
/translation="MSQTANDLRNNEVFFISPSNNTNKVLKDISQSEYKLVKNLUSGA
NQKRLIYDYNKQAYKIDMDNTSLITWNAPLSSVSVKTDNGDNQYVILQNTYISR
NVIIRTMYNLIQYIDIDTLMVTSQTSNNOFFKFSNICKTQISLTI
KFIDKONSNNVTIWSNNGDNQKWKILYNEKSMAYLTLCIKNNEVLTWFSSIGNVG
TYRTGNDQYWFVILNNDASMTYISNFSNQSKFLDVVNSGLADGTNVQVWDSNGTS
AOKWITPL"
4119..7709
/gene="ntaha"
4119..7709
/gene="ntaha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BAB71748.1"
/db_xref="GI:16580758"
/translation="MDINDDLNINSPVDKNVYIVRAKNTIFFKAKFVAPNIWAPE
RYIGEPDLIAEYKLDGDIYDSNLFQSDSERENFLQAIILKLRINNTISGKQLLSLI
STAIPFPGYIGGYSNPNIFTGKTPKSNKLSVSTIPFPFGYRETNYIESON
KNFYASNIYFPGSNIYVNNVYKKNDAENGMTAEIVFQPLTYVKNFYIDP
AMELTKGLISLFLYGLIKPSDLNLYPRLTDLNKPQSQLNIDLLISGGVDLEFI
NTNPTWNTSIFPNKMEFYKNIYKIEGNAINDKILKRLKQKQFQINQVDLWNL
NLNFYOSFNSIPDRFSNALKHFYKQYTYMDYTDYNINGFVNGQINTEKLESNKN
TNIISKEPVYVLYNNRNISLMKSNYDGLKGTEDFYSTYKIPYNEEYKRFNDSO
NPLNLSISIEVDSPHIIIDINPKNSDNLVFTQITSMTEVTTTALSINVLQAOI
TNNFVTLSDPSKSVSSKSLVYSGFLDNLMSYLETIKNDGPIDTKKYLWLKRVF
KNYSFDNLQOEIDSMOGINEVLMFKGKALNLTNSPVEEQDSGALISLKKDNL
REPNIEDIDSLSLGSFKDLANKLEYISKNIVYFKIYISFSLDQWTEYYSQYFE
LIMAKQSILIAQESLVKQIVONKPTDLSKASIPPTLKLIRETEKFTFDLNSQSIS
MNRVDNLNKAICFVEDIYKPIFYSEKYNINIKITREFRQCTNIDNEKLSILI
NSYFTKIDFKFLDIQSIKNFNSQVEQVMEKILSPYQLLLFASKGPNNSIEDISGK
NTLIQYTESIELVGVNGESLYLKSNETIKFSNKFTNGLTNNTICFWLFTGKND
DKTLIGNKVNCGWEIYFEDNGLVFEIIDSNGOVSYSVLSNIDSNVQYIISVDRL
KQDLLIRINDKNKVASIDOLLSYSTNITSLVKNKNSLYVEELSNDLPNTSEBIV
NYFSYLDNSYRDSKSLBYNNKYQLYNYVFPETSLYEVNDNNKSYLSLKNYDGINI
SSVAFKLNIDESYVYQKWEDECTICVLDGTEKYLIDISPNNRPIQLVSKDNKAKITI
NTDLFRPDCITFSYNDKXIFSLSLRDGDYNNMICNDNNKVPKGAHLWILES"
7724..11599
/gene="nt"

```

```

7724..11599
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/protein_id="BAB71749.1"
/db_xref="GI:16580759"
/translation="MPITINNYSDEPVKNKNIYLDTHLNTLANPEKAFRITGNIW
VIPDRFRNPNLNKPPRVTSPKSYDPNPYSTDSKDTLKEIKLFRKINSREI
GEELIYRLSDIIPPGNNTPIPTDFDVFNSVDVKTQGNWVKTGINSPIVITG
PRENIDPENSTPLKNNPFAOEGGALSILISPREMLTYSNANDVGEGRFSKSE
FCMDPILILHEENHANNENLYGAIANDQTSITTSNIYFSQYNKLEVAEYIARGGP
TIDIPKASFEERKALDYIRSIARLNSITTSANFNKNGYIGETQKILKIRKRVV
ESSEGVNKNFVELYNELTQITFEFNKAIYVQNKRIYLSNVTTPVTANILDDNV
YDQNGENIPKSNLNVLFMGONSRNPPALRVNPNENLYLTKFCHKAIDGRSLYNT
LDGRELIVKNTDFFIGDISDVKTDFLRKDINEETEVYYPDNVSDOVILSKNTSE
HGOLDLYPSIDSESEITLPGENOVFYDNRNTQNDVLYNSYVLESQKLSNVEDFTFR
SIEPALNSAKVYTYPTLANKVAGVGLFILMWANDVVEFTNLRKDLKDISD
VSAIPTYGPALNATISVREGNEFEAPVGYVTLLEAFPEFTIPALGAFVIYSKQVE
RNEIITKIDNLEQRKRWKDSYEMWGTWLSRIITQPNNISYQWYDSLNYAGAIIKA
KIDLEYKKSQSDKENIKSQVENKLSLDVKISEAMNNINKIRECSVTILFKNNLPK
VIDELNEFDRTAKLINLIDSHNIIILVGEVDKAKVNNSPQNTIPFNISYNNLS
LKDITINEFFNNINDISLQNRKNTLVDTSGTNAEVSQGVQNPFPDFKLGSS
GEDRGKIVYNNINIVNSYSEFSFWRINKVSNLPGYTIIDSVKNNESGWIGI
ISNPLVTLKQNSDSQSINFSDISNAPGVNKWFEVTVNMMGNMKIYINGKLID
TIKVKELTGTFESKTITPEINKIPDTGLITSQSDNIMKATRDYIFAKELQKDLNIL
FNSQYTNVYKDWGNDLRNKKEYVWVNDYLNRYMYANSROIYVNTNRNNDNFEGY
KILIKIRGNTINDIRVGGDILYFDMTINKKAYNLFMKNETMADNHSTEDIYALGLR
EQTKDINDNIIQIQPMNNTIYASQIFKSNFNGENISGICISIGTYRFRGLGDWYRHN
VLVPTVKQGNVASLLESTHGWGVPVSE"
BASE COUNT 4574 a 1263 c 1622 g 4288 t
ORIGIN

Alignment Scores:
Pred. No.: 11747
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
Dbs: 1 Gaps: 0

US-09-910-186a-10 (1-450) x AB061780 (1-11747)

Qy 2 ThrileProPheAsnIlePheSerTyrrAsnAsnSerLeuLeuLysAspIleLeuAsn 21
|||||
Db 10250 ACAATACCCCTTTATATTTTTCATATACATAATAATCTTTATTAAAGATATAATTAAT 10309

Qy 22 GluTyrrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
|||||
Db 10310 GAATATTTCAATAATTAATTAATGATTCAAAAATTTTGACCTACAAAAACAGAAAAATACT 10369

Qy 42 LeuValAspThrSerGlyTyrrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
|||||
Db 10370 TTAGTGGATACATCAGGATATAATCCAGAACTGAGTGAAGAGCGGATGTCAGCTTAAT 10429

Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
|||||
Db 10430 CCAATATTTCCATTTTGACTTTAAATTTAGGTAGTTACAGGGAGGATAGAGTAAGATTATA 10489

Qy 82 ValTrpGlnAsnGluAsnIleValTyrrAsnSerMetTyrrGluSerPheSerIleSerPhe 101
|||||
Db 10490 GTAACCCAGATGAATAATTTGTATATAATTTCTATGATGNAAGTTTACGATAGTTT 10549

Qy 102 TrpIleAa-gileAsnLysTrpValSerAsnLeuProGlyTyrrThrIleLeuAspSerVal 121
|||||
Db 10550 TGGATTAGATAATAATAATGGTAAATTAATTTACCTGGATATACATAAATGTAGTGT 10609

Qy 122 LysAsnAsnSerGlyTrpSerIleGlyIleLeuSerAsnPheLeuValPheThrLeuLys 141
|||||
Db 10610 AAAATAACTCAGGTGGAGTATAGGTATTATTAGTAAATTTTTTAGTATTACTTTAAA 10669

Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrrAspIleSerAsnAlaPro 161
|||||

```

```

1 (bases 1 to 12297)
Hauser,D., Eklund,M.W., Boquet,P. and Popoff,M.R.
Organization of the botulinum neurotoxin C1 gene and its associated
non-toxic protein genes in Clostridium botulinum C 468
Mol. Gen. Genet. 243 (5), 631-640 (1994)
94301293
8028579
2 (bases 1 to 12296)
Hauser,D.F.
Direct Submmission
Submitted (03-MAY-1993) D.F. Hauser, Inst. Pasteur, Unite des
Toxines Microbiennes, 28, Rue du Docteur Roux F-75724, Paris Cedex
15, FRANCE
On Jul 27, 1994 this sequence version replaced gi:509268.
Related sequences: X53751 & X66433.
Location/Qualifiers
1. 12297
/organism="Clostridium botulinum phase 1C"
/strain="468"
/sub_species="type C"
/db_xref="taxon:35348"
/lab_host="Clostridium botulinum"
289. 294
/note="ORF22"
302. 841
/note="ORF22"
/codon_start=1
/transl_table=11
/protein_id="CAA51308.1"
/db_xref="GI:509269"
/db_xref="SPTREMBL:Q38195"
/transl_table="MNDL5VATLENLKHDFNMFEMSLKKYIEKTSKKNLYLYDVYN
DILYHLWKELIELNLKFNSELDLRKYISTKRYCINICKKRNDRKKIIVNSEVYIK
KLDAVNYSLYCNFEEFLDLISILNKEQIYMKFTEGRKDNEIARLRLSRQSIY
IRITSKLKPIVMQVLNI"
gene
complement(854. 2787)
/gene="ANTP-70"
854. 906
terminator
/note="putative; ORF22"
complement(854. 906)
/gene="ANTP-70"
/gene="ANTP-70"
/note="putative"
complement(916. 2787)
/gene="ANTP-70"
/codon_start=1
/transl_table=11
/protein_id="CAa51309.1"
/db_xref="GI:509270"
/db_xref="SWISS-PROT:P46085"
/transl_table="MELSKELYITKDSINNVLADGNVYVNRGDWILSRQNQLG
FSTNNGCTAGDGLREITATPYYPYTPASNEEYIKNNQNVFANFEASEIPIGE
FKSTAFSNKSLMYLQYTVIRYEIIKVLQNTVTRAVLYVPSLGVKSIFFNSEEQID
KNYFVSQDKILNEKFIYVKYDDTIYKESKNNNINFNFTSOTILPYNPGYLVYN
GDGMYRNDKDLGLTLLIESSTSGSIIOPRLNTRPLFNSTPTIFSQVTEARLND
AFNIQLNPTSLFKVEEAPNPKISMKYNTVEKYLINXONGNIDDKAEYLLPSL
GKEVSDASQPQAVYETVDDQGIQTGNENIIVGVINPSENIEISSTPIEDDITY
NIPTSIQNNACVYLFVKNTYGYKTTKNNPLPLIIYEAIGSSNRNNNSNNLSNDNIK
AIKYITGLNRSKSAKSLYLSLPGKNRYIRIPOPSSSTISQLIFRFLGHNISLDAST
VNIILNDTSGTHYRQSPDQNTISYQLTIPGDFNNIASSIFSFTRNNQIGILY
RLTESINGYLNITPNNSDLLNNVEPISLLNGATVIFRVKVTLENNYNIIFDAYRNS"
complement(2801. 3252)
/gene="ANTP-17"
complement(2801. 3241)
/gene="ANTP-17"
/codon_start=1
/transl_table=11
/protein_id="CAA51310.1"
/db_xref="GI:516172"
/db_xref="SWISS-PROT:P46083"
/transl_table="MSERTIFPNNGNYIKSLFSDSLYLYTVSSGALSFSTSSLDNOK
WKLEYTSSNGFFSFAENPKNFLAYNDYGFYILSSSSNNSLNWPKIRAINSYIICIT
SYNVTDYANTYIDNNNNITDQPIILNPNFDINNSNQILKLEK"
complement(3247. 3252)
/gene
CDS

```



```
gene
complement(3304. .4176)
CDS
/gene="ANTP-33"
complement(3304. .4164)
/gene="ANTP-33"
/codon_start=1
/transl_table=11
/protein_id="CAA51311.1"
/db_xref="GI:516173"
/db_xref="SPTREMBL:Q38196"
/translation="MSQTANDLRNNEVFVSPSNNNTKVLKDISKEVKNLKLSSGA
NOKRLAYDTNKKYKIKVMDNTSLITWNAPLSSVSKVDTNGDQYKYLQNTLSR
NVIIRNPNLVLQVNDITDMLAYSTQSSNOFFKFSNCIYEALNRACKLQTLNS
DRFLSKLNSOILVLMQWFOSSROKWIIEYNKSAVTLKCOENNRYLTIWQNSNVY
ETVOSTSLIQWNIWLNLDNDASKYIILYNQDINRVLVDVNSQIANGTHVIVDSYHN
TNQKWLINLI"
complement(4171. .4176)
RBS
/gene="ANTP-33"
4313. .8017
/gene="ANTP-139"
4313. .>8017
RBS
4411. .4416
/gene="ANTP-139"
4427. .8017
CDS
/gene="ANTP-139"
/codon_start=1
/transl_table=11
/protein_id="CAA51312.1"
/db_xref="GI:516174"
/db_xref="SPTREMBL:Q38197"
/translation="MDINDDLNINSPVDNKNVIVRARKNTNFFKAFKVPAPNIWAPE
RYGEPLDIAEYKLDGDIYDSNPLSQDSERENFLQAIILLKRINNTISGKQLLSLI
STAIPIFYGGGYSNPNIPTFGKTKSNKLNLSVTSTIPPPFGYRETYIEQSN
KNPYASVNIIFGGSNVNNVYKKNDAENGMTMAEIVFOPLLTYKYNKYIDP
AMELTKLILGLYGLKPSDLNVPYRLTELDAKQESOLANIIDLSSGGVDLEFI
NNTWYFNSVEPNSIKMFYKKNVYKTELEGNNAIGNDIKPLKQKFOINQVDIWNL
NLNFCQGSNIIIPDRESNALKHFKRYQYTYNDYTNNYINGFVQGINIKPLPSKN
NIIISKPVVNVNENISLMKSNIIYDGLKGTTFEDYSTYKIPNEYESYRFNDSQ
NPPNNSIEVDSPIEIIIDINPKNSDLNFTQTSMTTEVTTTALSINYLQAI
TNNEFTLSDFSVKSKSLVYSFLDNLSYLETIKNDGPIDTDKRYIWLKBEV
KNSYFDNLTOEIMSGCINEVLFWGLKALNLTNSFVEEYQDSGALSLSKKNL
REPNIEDLSDLSGLKDLNKNLYEIKSVIVFKIYFSLFQWTEYYSOYFE
LICAKQSIAGSLVQIQVQNKFTDLKASIPDPLKILRETTKTFIDLNSQIS
MNVVDNFKASICVYVEDIYKPEISYKRYINNINIKREFIQRCTNINDNEKSILI
NSYFTIDFLDIQSIKFFNSQVEQWKEILSPQLLLFASKGPNSEILEDLSGK
NTLIQYIESILVYGVGESLYLKSPPETIKFSNFTNGLTNFTICFWLRTGKND
DKRLIGNKVNCCWEIYFEDNGLVFEIIDSNGQESVYLSNINDNWTYIISTVDRL
KDQLLIFTNKNVANVSIDQILSYSTNIIISLVNKNRSIYVEELSVDNPIYSEVIR
NFFYLDNSYTRDSKSLLEYKNQYNYVFPETSLYEVDNKNKSYLSLKNTDGINI
SSVKFLNIDESKVVQKQDECICVLQDTEKYLDISPENNRIQLVSSKDNKAKTIV
NTDLPRPCITPFSNDXYFSLSLRDGQENMIDNDNNKVPKGAHLILES"
7931. .12032
/gene="BONT/C1"
7931. .>12032
RBS
/gene="BONT/C1"
8016. .8021
/gene="BONT/C1"
8032. .11907
CDS
/gene="BONT/C1"
/codon_start=1
/transl_table=11
/protein_id="CAA51313.1"
/db_xref="GI:516175"
/db_xref="SWISS-PROT:P18640"
/translation="MPTINNFKYSDPDVKNKILYLDTHLNTLANEPKAFRTGNW
VIPDRSRNPNLKNPRTVSPKSGYDENVLSTDSDDTFLKELIKFKRLNSREI
GEELIYRUSTDIPPGNNPTPIITFDVDFNSVDVKTROGNWNVKIGSINSEIIT
PRENIDPSTFKLTNTNTAAQGFGLSIIISPRFMTYSNATNDVGEGRFSKE
FCMDPILTLMEHLAHMHNLYGAIAPNDQIISVTSNIFYSYNVNKLVEIYAFQSG
TIDLTPKARKYFEEKALDYIRSIKRLNSITNAPNSFNKYIGEYKQKILIRYRVV
ESSEGTVNRNKFVELYNELTOIFTEFNKAIYNQVNRKLYLSNVYTPYNTLDDNV
YDIQNGFNIKPSNLNLFEMQNLNRNPKVAPENKMLYLFTEFCHADIGRSLYNKT
```

```
LDCELLLVKNTDLPFICGDISDKVTDIFLRKDINEETEVIYVDPNVSDVILSKTSE
HGQDLDLPSIDSESEILPGENVFVNDNRQNTQNDVLYNLSQSLNSVDVDFPTR
SIEBALDNSAKVITYFTFLANKVAGVGGFLMWANDVVEDETTILRRDLDKISD
VSAIPIYIPALNISNSVRNGNTEAFATVGTGVTLLAEPEFTIPALGAFVYSKVOE
RNEIKITDNCLEQIRKWKDSVWMMGTWSRIITQFNNSIYQMTDSINYAGAILKA
KIDLEYKRYSGSKENIKSOVENLKLKGVNLDVLSIEMANNINKFIRECSVTYLFKNMLPK
VIDELNEFDNRNKAFLNLDISHNIIIVGVDKILKAKVNSFQNTIPFNIFSYNNSL
LKDIINEYFNINDSKILSLQNRKNTLADTSGYNAEVSSEBGDVLAPIPDFPKLSS
GEDRGKVIYQCNENIYVNSYSEFSISFWIRINKWYSNLPGYIILISVYNNSSWISGI
ISNLEVLITQCNEDSEQISNFSISNNAFGINKWFEFTVTNNMGNMKIYIINGKLID
TIKYKTEGLNFKSITITFEINKIPDGTGLITSDSNINMWIRDFYIFAKELDGGKIDIL
FNSLOYTNVNDYKWNDRYKXYVYVNDIYLNRYMYANSRQIVFTIRNNNDNFEGY
KIIILKIRGTNTDTRVGGDILYFDITNNKAYNLPMKNETMADHSEDYIAGILR
BOTKDINDNIITFOIQPNNTYVYASQIFKSNFNGENISGICSGITGYRFRXLGGSWRHN
YLVPTVQGNVYASLLESTSHWGFVPVSE"
11993. .12032
terminator
/gene="BONT/C1"
BASE COUNT 4825 a 1287 c 1703 g 4482 t
ORIGIN
Alignment Scores:
Pred. No.: 9,23e-152 Length: 12297
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 7 Gaps: 0
US-09-910-186a-10 (1-450) x CBCBONT (1-12297)
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 10558 ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTTATTAAGGATATAATTAAT 10617
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 10618 GAATATTTCAATAATTAATTAATGATTCAAAATTTTGACCTACAAACAGAAAAAACT 10677
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 10678 TTAGTGGATACATCAGGATATAATCAGAAAGTGAGTGAAGAAGCGATGTTACGCTTAAT 10737
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 10738 CCAATATTTCCATTGACTTAAATTAGGTAGTACGGGAGGATAGAGTAAGATTATA 10797
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 10798 GTAAACCCAGAAAGAAAATTTGTATATAATCTATGATGAAGTTTTCAGCATTAGTTT 10857
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuPheGlyTyrThrIleIleAspSerVal 121
Db 10858 TGGATTAGATAATAATAATGGTAAAGTAATTTACTGGATATACTATAATGATGATGTT 10917
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 10918 AAAATAACTCAGGTGGAGTATAGGTATATTAGTAATTTTTCAGTATTACTTTTAAAA 10977
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161
Db 10978 CAAATAGAGATAGTAGGACAAAGTATAAATTTAGTTAGTATATATAATAATGATGCTCT 11037
QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 11038 GGATACATAAATGCTTTTTTGTAACTCTTACTAACATAATGATGGGAAATATCAAGATT 11097
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 11098 TATAATAATGGAAAATTAATAGACTACTATAAAGTTAAAGAACTAACTGGAATTAATTT 11157
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 11158 ACCAAACTACTAACATTGGAATAATAATAATCCAGATACCGGCTTTGATTACTTCAGAT 11217
```

QY 222 SerAspAsnIleAsnMetTTPlleArgAspPheTyrllePheAlaIysGluLeuaspGly 241  
 |||||  
 Db 11218 TCTGATACATCAATATGTGATAAGAGATTTTATATATTTGCTTAAGAATATAGTGGT 11277  
 |||||  
 QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValIalysAspTyrtip 261  
 |||||  
 Db 11278 AAGATATATATATATATATATAGCTTCAATATACATAATGTTGTAAGAATATTGG 11337  
 |||||  
 QY 262 GlyAsnAspLeuArgTyAsnLysGluTyTyMetValAsnIleAspTyrtipLeuAsnArg 281  
 |||||  
 Db 11338 GGAATGATTTAAGATATAATAAAGATATATATATGTTTAATATAGATATTATAATAGA 11397  
 |||||  
 QY 282 TyrMetTyAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
 |||||  
 Db 11398 TATATGATGCGACTCAGCAAAATGTTTAAACACAGTAGAATATATATGACTTC 11457  
 |||||  
 QY 302 AsnGluGlyTyrlsIleIleIleArgGlyAsnThrAsnAspThrArgVal 321  
 |||||  
 Db 11458 AATGAAGGATATAAATATATAAAGAAATCAGAGGAATACAAATGATACATAGAGTA 11517  
 |||||  
 QY 322 ArgGlyGlyAspIleLeuTyPheAspMetThrIleAsnAsnLysAlaTyAsnLeuPhe 341  
 |||||  
 Db 11518 CGAGAGAGAGATATTTATATTTGATATGACATTAATACAAAGCATATATTTGTT 11577  
 |||||  
 QY 342 MetLysAsnGluThrMetTyAlaAspAsnHisSerThrGluAspIleTyAlaIleGly 361  
 |||||  
 Db 11578 ATGAAGATGAACATGTATGTCAGATAATCATAGTACTGAAGATATATATGCTATAGT 11637  
 |||||  
 QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
 |||||  
 Db 11638 TTAAGAGAACAAACAAGGATATAAGATATAATATATATTTCAATACACCAATGAAT 11697  
 |||||  
 QY 382 AsnThrTyTyTyAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
 |||||  
 Db 11698 AATACCTATATATAGCATCTCAATATTTAAATCAAAATTTTAAATGAGAGAAATATTC 11757  
 |||||  
 QY 402 GlyIleCysSerIleGlyThrTyArgPheArgLeuGlyGlyAspTyrtipTyArgHisAsn 421  
 |||||  
 Db 11758 GGAATATGTTCAATAGGATCTATCGTTTAGCTTGAGCTGGAGTGATGCTATAGACACAAT 11817  
 |||||  
 QY 422 TyrLeuValProThrValLysGlnGlyAsnTyAlaSerLeuLeuGluSerThrSerThr 441  
 |||||  
 Db 11818 TATTTGGTGGCTACTGTGACGACAGGAATATATGCTTCATATTAGAAATCAACATCACT 11877  
 |||||  
 QY 442 HisTrpGlyPheValProValSerGlu 450  
 |||||  
 Db 11878 CATTTGGGCTTTGTACCTGTAAGTGAA 11904  
 |||||  
 RESULT 8  
 LOCUS AB037920 11631 bp DNA linear BCT 24-JAN-2002  
 DEFINITION Clostridium botulinum genes for ORF-22, HA-70, HA-33, NTNHA,  
 neurotoxin, complete cds.  
 ACCESSION AB037920  
 VERSION AB037920.1 GI:6939789  
 KEYWORDS neurotoxin; NTNHA; HA-33; HA-17; HA-70; ORF-22.  
 SOURCE Clostridium botulinum (strain:D-4947) DNA.  
 ORGANISM Clostridium botulinum  
 Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 Clostridium.  
 REFERENCE 1  
 AUTHORS Kouguchi,H., Watanabe,T., Sagane,Y., Sunagawa,H. and Ohyama,T.  
 TITLE In vitro reconstruction of the Clostridium botulinum type D  
 progenitor toxin  
 JOURNAL J. Biol. Chem. 277 (4), 2650-2656 (2002)  
 MEDLINE 21659747  
 REFERENCE 2 (bases 1 to 11631)  
 AUTHORS Sagane,Y., Watanabe,T., Kouguchi,H., Yamamoto,T., Takizawa,J.,  
 Kawabe,T., Murakami,F., Muroga,A., Nakatsuka,M. and Ohyama,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-FEB-2000) Tohru Ohyama, Tokyo University of  
 Agriculture, Faculty of Bioindustry, Yasaka 196, Abashiri, Hokkaido

099-2493, Japan (E-mail:t-oyama@bioindustry.nodai.ac.jp,  
 Tel:81-152-48-3838(ex.368), Fax:81-152-48-2340)  
 Location/Qualifiers  
 1. .11631  
 /organism="Clostridium botulinum"  
 /strain="D-4947"  
 /db\_xref="taxon:1491"  
 6. .545  
 /gene="orf-22"  
 6. .545  
 /genes="orf-22"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ORF-22"  
 /protein\_id="BAA90656.1"  
 /db\_xref="GI:6939790"  
 /translation="MNDLFYATENLKHONHDFLEMSLKVKYIEKTSKKYNLYDYVN  
 DILHMKELIENKENSELDLKYISTSKRYCINICKKKKDKKLIYNSATYK  
 KLEAVNYSLYCEDEFELDLISILYKKEQIIYMAFFECRKNDEIARHLHLSRSIYK  
 IRIKSLKLYPIVMQLVNI"  
 complement(621. .2492)  
 /gene="ha-70"  
 complement(621. .2492)  
 /gene="ha-70"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-70"  
 /protein\_id="BAA90657.1"  
 /db\_xref="GI:6939791"  
 /translation="MSLSIKELYTKDKSINNVLADGNVYVNRGDGWLRSQNQLG  
 GNTSNGCTAIVGDLRIETATPYYPPTASENEEYIRNNVONVPANFTEASEIGIGE  
 FSKTASNGKLYMYLOYTYIRYELIKVLNTVIERAVLYVPSLGVAKSIENSQEQID  
 KNFYSEDKCLNERFIKKIAETIKAKESNDNSNTNLNTSQTILPYPNGLYVINK  
 GDGYMTNDKDLIGTLLEIETNSGSIQPLRNTTRPLFNTSNPLFSCETEAKLND  
 AFNIQNTSTLTFKEVEAPDNKISMKAYTYEKELINYQNGNIADKAEYLPSL  
 GKEVDASPPQAVETVEQGTQGPENIIVGVINSENTEESTIPDDTY  
 NITPSONNACVLFVTNTGVYKINAONNPLIYIESIGSDNNIQTSLNNNIK  
 AINWGTSSNAESVLYSLIKKNYIRIPQISSSTNOLIFKRELGNISDLNST  
 VLINLNTSGHYHTROSPDYGNKISYOLTIPTGDNNTASSIFSFRNNQIGTLY  
 RLDESINGNLIRIKNISDLNNEPISLNGATYIFRVKYTELNANYIIFDAYENS"  
 complement(2506. .2946)  
 /gene="ha-17"  
 complement(2506. .2946)  
 /gene="ha-17"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-17"  
 /protein\_id="BAA90658.1"  
 /db\_xref="GI:6939792"  
 /translation="MSSERTFLPNNGYKIKSLFSDLSLYTSSGSLSTANTSSLDNQK  
 WKLEYISSNGPRFSNVAPNPKYLAYNDYGFYLSSSNNSLWNPDKIAINSYIICTL  
 SIYVTDYATVIYDNNNTIDQILNLPNFDINNSQILKLEK"  
 complement(3008. .3868)  
 /gene="ha-33"  
 complement(3008. .3868)  
 /gene="ha-33"  
 /codon\_start=1  
 /transl\_table=11  
 /product="HA-33"  
 /protein\_id="BAA90659.1"  
 /db\_xref="GI:6939793"  
 /translation="MSQTANDLRNNEVEFFSPSNTNKVLKISQSEVKLWKLKLSGA  
 NKWRLYTDNKQATYKVDNLTSLITWDAPLSSVSVKTDNTNNQWYLQDYISR  
 NVILRNMPNLVQYVTDITLIVSTQTNNSQFFKSCNICYEALNNRCKLQTLNS  
 DRFLSKNLNSQIIVLQWFDSSRQKWTIEYTKSAYTLKCOENRYTLTQNSNRY  
 ETQVSDSLIQYWNINYLNDASKYILYNLQDTRVLDVYNSQTANGTHVIVDSYHG  
 TNOQWIINLI"  
 4129. .7719  
 /gene="ntnha"  
 4129. .7719  
 /gene="ntnha"  
 /codon\_start=1  
 /transl\_table=11

```

/product="NTNIA"
/db_xref="GI:6939794"
/translation="MDINDLNINSPVDNKNVIVRARKINTPEKFAKVPNIWVAPE
RYGEPLDIAEYKLDGDIYDSNPLSODSRENFLOAIITLKLINNTISGKQLLSLI
STALPEYGVGGYSSNPIFTGKTPKSNKLSNLSVTSTPPFGGIRETNINISQN
KNKYASNIIVFGSGNIVENNVCYKKNDAENGMTMAEILFPQLPILTYKNFYIDP
AMELTKCLIKSLFLYIGLKDSDLLVPYRLRSTLDNKPQSLNIDILLISGGVDEFI
NTNPTWTSNYSFISIMFKYNIETIEGNAIGNDKLRKQKQFNSQVQDIWNK
NLNFTSFENSIIIDPFSNALKHFRKYQTYTDGNYINGFVNGQINIKLPLSDKN
TNIISKQEVLYGVNGESLYLKSPNETVEFNFTGLTNFTICFWLFTGRDD
DKRLIGNKVNCCWEIYFEDNGLVPFELIDSGNQBVSYLNNVNNKYISISVDRL
KQOILLIFINDKRNVAIEQILNIYSTNVLVLYNKNNSIYVEELSVDLPVASEVIR
NYFSYLDNYSIKSLLSEFINKNYQINIKYFETSLYEVDNKNKSLKNTDGINI
PSVAFKLINIDESGYVQKDECLICVSDGTERYLIDSPENNRIQLVSKDKNAKITV
NTDLFRPDCITFSYNDKYFSLSRDSDYNMCMNDNNKVPKGAHLILKS"
/734..11591
/gene="nt"
/734..11591
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/db_xref="GI:6939795"
/translation="MTWPVDFNYSQVSDVNDILYLIPONKLIITPVKAFMTNRDIW
VIPRFSDDTNPISLKPRTSPKSYDPSYLSLSTDEQDTFLKGIILKPIKNERDI
GKLLNLYVGSPPKSSPEPTEDTRHTINAVEKEPENGSKWYNITISVYLFG
PLPNIIDTASLTQGOQSPSEFGTILSLVAPEFLLTFSDVTSNOSSAVLGKSI
FCMPDVALMHETLSHQYIGINIPDKIRPQVSEGFSDQPNQVEEYTFGGS
DVEIIPQRLQREKALGHYKOTAKELNNIKTIPSSWSNDIKYKIFSEKYNPKD
DNTGNVYNDIKFSLSDTNNVSEVYYSQYNKRTHYFSKHLYPVFANILDDNI
YTIINGNLTFTKFNIGNSQNIERNPALOKLSSESVDLFTKVCRLTRNRDSDTC
IQVKNLTYPVADKDSI:SORIFESQIITDETNVNTSDNLSDELDAKVPNTPEAV
DPLLPVNMPELVPNGSEEVYDDITKDVYLSNXYLEAKLNNVNITLTSVEE
ALGSNKYITFLPSLAKYKNGVQAGFLKAWANEVVEDTITNMKDDTLKSDVSAI
IPYIGALNINISALRNGKQAFATAGVAFLEGFPEFTIPALPGVTFYSSIQERKI
IKTENCLEQVRKWKDSQYMWNSWLSR:ITRNHISYQMSDLSYQADAKKIDL
EYKSGSKENIKSQVENLKNLSDYKISEAMNNINKFIRECSVYLFKMLPKVYIDE
LNKPDKTKTILNIDSHNILLVGEVDRKAKVNESFENTIPFNIFYSTNLSLLKDM
INXFNSTNDSKILSLONKNTLMDTSYNAEVRVEGNVOLNIPFPDFKLGSSGDR
GKVIATONENIYVNAVYESISFWIRINKVSNLPXYTIIDSVKNKSGWSIGLISF
LVFLTKQNESEQINISYDISKNAAGYKWFVITITNMGNMNIYINGKLIDTIKY
KELTGINFSTIITOMKNIPNTGLTSDSNINWINDFYIFAKELDKDDKINILFNSL
QTNVVDKWDYNGDKKEYYMINVNMNRYMSKKGNGIVFNTRKNNDFNEGYKII
KRIGNTDTRVGENVLYFNTITDNNKQYSLGMYKPSRNGLDVLPLGLDQPMDEIR
KYGGFIIQPCNTFDYASQLFLSSNATNRLGILSGYSFKLGGDDYWFNHEYLIPVI
KIERVASLESTSTHWVFPASE"

```

BASE COUNT 4493 a 1254 c 1657 g 4227 t

ORIGIN

Alignment Scores:  
 Pred. No.: 7,46e-114 Length: 11631  
 Score: 1824.00 Matches: 352  
 Percent Similarity: 85.84% Conservative: 36  
 Best Local Similarity: 77.88% Mismatches: 56  
 Query Match: 76.25% Indels: 8  
 DB: 1 Gaps: 4

US-09-910-186a-10 (1-450) x AB037920 (1-11631)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
 Db 10248 ACATACCCCTTTAAATATTTTTCATATACTATAATATCTTTTATAAAGATATGATTAAT 10307

QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 Db 10308 GAATATTTCAATAGTATTAAATGATTTCAAAAATTTTACCTTACAAAATAAAAAATAACT 10367  
 QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 Db 10369 TTGATGGGATACATAGGATATAACCCAGAGTAGAGTAGAAGCAATGTTTCAGCTTAAT 10427  
 QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 Db 10428 CCAATATTTCATTTGACTTTAAATAGTAGTATTCAGGGGATGATAGAGGTAAGTTATA 10487  
 QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
 Db 10488 GTAAACCCAGATGAAAAATTTGATATATGCTATGATGAAAGTTTGTAGTATTAGTTTT 10547  
 QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
 Db 10548 TGGATTAGGATAAATAAATGGTAAAGTAATTTACCTGGATATACATAATGATAGTGT 10607  
 QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
 Db 10608 AAAATAAATCAGGTGGAGTATAGGTATTTATGATATTTTGTAGTGTACTTTTAAAA 10667  
 QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161  
 Db 10668 CAAATGAAATAGTGAACAAGATATAAATTTAGTTATGATATATCAAGAATGCTGCG 10727  
 QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
 Db 10728 GGATATAATAATGTTTTTTGTACTATTACTACCAATATGATGGGAATATGATGATT 10787  
 QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
 Db 10788 TATATAATGGAAATTAATAGATACTATAAAGTTAAAGAGTTAACTGGAATTAATTT 10847  
 QY 202 SerLysThrIleThrPheGluIleAsnLysLysIleProAspThrGlyLeuIleThrSerAsp 221  
 Db 10848 AGCAAAATATACATTTTCAATGAATGAATTTCCAAATACCTGCTTAATTAATCCTCAGAT 10907  
 QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
 Db 10908 TCTGATACATCAATATGCGGATAGGATTTTATATCTTTGCTAAGAATTAGATGAT 10967  
 QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrp 261  
 Db 10968 AAAGATATTAATATATTTTAAATAGCTTGCATATACTAATAATGTTGTAAGATTTATGG 11027  
 QY 262 GlyAsnAspLeuArgTyrAsnLysGlnTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
 Db 11028 GGAATGATTTAAGATATGATAAAGAATATTACATGATTACGTAATATTATATGAATAGA 11087  
 QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
 Db 11088 TATATGCTTAAAAAGGCAATGGAATTTGTTTTAATACACGTAATAATAATAATAGCTTC 11147  
 QY 302 AsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
 Db 11148 AATGAAGGATATAAATATATAATAAAGAATTTAGAGGAATACAAATGATACATAGAGTA 11207  
 QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
 Db 11208 CGAGGAGAAATGTTATTTATTTAATACTACAAATGATAACAACAATAATAGTTTA--- 11264  
 QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThr---GluAspIleTyrAlaIle 360  
 Db 11265 -----GGTATGATATAAACCTTCTAGAAATPAGGACTGATTTAGTTCCTACTA 11312  
 QY 361 GlyLeuArgGluGlnThrLysAsp---IleAsnAspAsnIleIlePheGlnIleGlnPro 379  
 Db 11313 GTGCGATTCGATCAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11372  
 QY 380 MetAsnAsnThrTyrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsn 399







```

NVLIRNPNLVLYQWNTDITLVTSQTSNNQFFKFSNCIYEALNNRNCKLQIOLNS
DRFLSKLNQSIIVLWQWFSRQKWLIEYNETKSAITLKQENNRNYLTWQNSNYV
EYQTSQSLQYNNLYNDNDASKYILYNLQDTRVLDVYNSQIANGTNNVIVDSYHG
TNOQWIINLI"
4145..7735
/gene="ntaha"
4145..7735
/gene="ntaha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BA089712.1"
/db_xref="GI:6729212"
/translation="MDINDOLNINSPVDNKNVIVVARKTNTFFRAFKVAPNWIATPE
RYGEPDLIAEYKLDGDIYDNLPSODSERENFLQAIITLLKRNINISGKQLLSLI
STALPIYGVGGYSNPFTFKTKPKSKNKLNSLVTFITPFPFGYRETNYSIPSN
NKDYAFNIVFGPNSVNNVYCYKKNKAENGMTMAEILFQPLLYATYKNKFIIDP
AMELTKGLISLFLYGIKPSGLVYVRLRLTELDNKPQSLNIDLLISQSGDLEFI
NTNPTWNTSVFNSIKMFKYKNIYETEIEGNAJNDIKRLRQKFGNSQVDIWNL
NLNYSKEPNSIIPDRESNALKHFKYKQYTMDDYDNYNGINGVNGOINIKLPLSDKN
TNLISKEKVVNLVNNNISMKSNIYDGLKGTEDFYSTYKIPNNEYEYRPNDS
NPLNNISIEVDIPETIIDINIKONSDNLVETQITSMTEEVYTHRALPYNLQAI
TNSFNFLSDFSKVSCKDSLYVSEFLDNLSYLETIKNDGPDIDDKKYLWLKEVF
KNYSFDNLQEDSSGKNEVWFGKALNINLTNSFVEYQNSGPISLSKKNDL
SEPIEIDIDPSLGLSFSKDLANKLYEYISKNIYFKIYFNFLDQWTEYYSQYFE
LICMAKOSILQAESLVCIIQNKFTDLSKASIPDYLKIKETKFTFDLSKESQIS
MRYVDNFKASICVVEDIYPKFISYMEKYINNINIKUREPQRTQNTNDEKSLI
NSYFKPDLKFDLQKIKFNSQVQWKMELSPKQLLLFATRGPSNLIIDTSGK
NTLQYVESVLYGVGNSLSLYLSPNEIVFESNFTNGLNNFTICFWLFTQKDD
KDLILGNKNVKNQWEIYFDNGLVFIIDNSGNQSVLSVNNNNWYIYISVDRS
KQDLIFNKNVKNQWEIYFDNGLVFIIDNSGNQSVLSVNNNNWYIYISVDRS
NFGYLDNINDESQSLSKLYNKNYQLYNVPFKTSLEYEDNNKSYLSKNDGINI
PSVFKLINDIESQYVQKWDCEIICVDSGTKEYLDISENNRQIOLVSKNNAKITV
NTDLFRPDCITFSYNDXYSLSLRDGDYNNMINDNNKVPKGAHLILES"
7750..11592
/gene="nt"
7750..11592
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/protein_id="BA089713.1"
/db_xref="GI:6729213"
/translation="MPITINNFNVSQVDNKNLYLDTHNTLANPEKAFRIIGNIW
VIPFRERDSNPNLKNPRTYSPKSGYDNPYLTSDSEKDTFLKEIKLFRKINSREI
GELIYRLATIDPPGNNNTPIINFDFDVFNSVDVKTROGNNWVATGINSVITIG
PRENLIDETSTPKLTNTAAQDFGALSIIISIPREMLTYSNATNNVGEGRFSKSE
FCMDPILILHMLHAMNLYGTAIPNDQRISSVTNIFYSOYKVKLEAYIYAGGP
TIDLIPKARKYFEKALDYRSIAKRLNITITANPSSFNKYIGEYKQKILKRYEV
ESGEVADRKKFELIKELTQITFEENAKIYNQNRKIYLSNVTIPYANILDDNV
YDIQNGFNIPKSNLNFQGNLSRNPALRKVPENMLYLFTKFKHAIIDGRSLYNTK
LDCEILVKNLDFIDIGDISDKTDFIDINEETEVIDYDPNVQDVQVILKNTSE
HQDLDLPIIIEGSQLVPGENQVYDNRQNDVYLSNYLYLESQKLSDNVEDFTFTT
SIEALDNSGVYTYFFPKLADKVNVTQGGGLFLMWANDVVEFTTILKRDITDKISD
VSATIPYIGPALNINSVRRCNTEPAFVGTWILLFAPOEFTIPALGAFVIYSKVOE
RNEILKTIDNLEORIKWDSYEMWIGTWLRLITQFNINISYQWDSLYNQADALFO
KIDLEYKISGDKENISQVENKNSLDIKISEAMNNINKFPRESVYILFKMLPK
VIDENLPDLKNTLINLIDSHNIIIVGEVDRKAKVNFESFENTIPFNITKNSL
LKDIINEFYINDSKILSQNNKALVDYSGVNAEVRLEDQVQVNTIYNFKNSLS
GDKIIVNNNNILYSAIYENSVSFWIKISDNTNSHNEYTIINSKQSGWKLCIRN
GNIWILQIDINRKYSLIFDYSESHTGTNKFVFTITNNIMYKLYINGELQOS
FRIEDNLVKDKITVFGIDENIDENOMLWIDROFNIFSKELSNEDINIVYEGOILNV
IKDWGNPLKDETEYIINDYIDRYAPKSNITLVLOYDPDRSKLVTGPIFKSVSD
ANPYSJUNGONMFMHMLNSGKWTIIRDDITAIIEGRCSKNCNVALKLQSNLNGY
GIGIFSIKIVSQNKYSQIFSPFMKNTMLLADIYKPRFSEFNAYTPVAVTNYETKL
LSTSFWKFIQRDPGWVE"
BASE COUNT 4499 a 1240 c 1651 g 4224 t
ORIGIN

```

Alignment Scores:  
 Pred. No.: 2,618-50 Length: 11614  
 Score: 882.50 Matches: 198  
 Percent Similarity: 61.20% Conservative: 78

```

Best Local Similarity: 43.90% Mismatches: 148
Query Match: 36.91% Indels: 27
DB: 1 Gaps: 10
US-09-910-186A-10 (1-450) x AB037166 (1-11614)
QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21
DB 10276 ACAATACCCCTTTAATATATTTTTCATATACATAAATCTTTTATTAAAGATATAATTAAT 10335
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAspThr 41
DB 10336 GAATATTTCAATAGTATTAAATGATTCAAAATTTTGAGCTTACAAAACAAAAAATGCT 10395
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGluValAspValGlnLeuAsn 61
DB 10396 TTAGTGGATACATCAGGATATAATGCAAGAGTGGAGGTAGAGGTGATGTTCAAGTAAAT 10455
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
DB 10456 ACGATATATACAAATGATTTTAAATTAAGTAGTTCAGAGAT-----AAAATTATA 10506
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
DB 10507 GTAAATTTAAATAATAATATTTTATATAGCCCTATTTATGAGAACTCTAGTGTAGTTT 10566
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120
DB 10567 TCGATTAGATATCTAAGATTTAACTAATTTCTCATATGAATATACATAAATTAATAGT 10626
QY 121 ValLysAsnAsnSerGlyTyrTrpSerIleGlyIleLeuSerAsnPheLeuValPheThrLeu 140
DB 10627 ATAAAAACAAATTTCTGGTGGAAATATGATTTAGGAATGCAATATAGAAATGCAATTTTA 10686
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
DB 10687 CAGATATTAATAGAAAATATAAAGTTTAAATTTTGATTAAGTGAATCATTAAGTCAT 10746
QY 161 ProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179
DB 10747 ACAGGATATACAAATAAATGTTTGTCTACTATACTAATAATAATATGCGGTATATG 10806
QY 180 LysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyLe 199
DB 10807 AAACCTTTTATAAATGGAGATTAAGCAGAGTGAAGATTTGAAGATTTAAATGAGCTT 10866
QY 200 AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219
DB 10867 AAGTTAGATAAAACCATAGTATTGGATAGATGAGAAATATAGAT----- 10911
QY 220 SerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeu 239
DB 10912 -----GAGAATCAGATGCTTTGGATTAGAGATTTTAAATTTTCTAAAAGAATTA 10962
QY 240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAsp 259
DB 10963 AGCAATGAGATATTAAATATCTATATGAGGACAAATATTAAAGAAATGTTATTAAAGAT 11022
QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
DB 11023 TATTGGGAAATCCTTTGAAAGTTGATACAGAATATTATATTAATTAATGATAATATATA 11082
QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299
DB 11083 GATAGGTATAGACCTTAAAGATAATATCTTGTCTGTCTAGTATCCAGATAGATCT 11142
QY 300 AspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAscThrAsnAspThr 319
DB 11143 AAATTATATCTGGAAATCTTATTACTATTAAATCAGTATCTGATAAGAATCCCTATAGT 11202
QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
DB 11203 AGAATTTTAAATGGAGATAAATAATATGTTTCATATGTTATATAATAGTGGGAAATATATG 11262

```



```

QY 340 LeuPheMetLysAsnGluThrMetTy:Ala-----AspAsnHisSerThrGluAsp 356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11263 ATAATAAGAGATAGTACATAATATGCAATAGAGAGAGAGTGTCAAAAAATGTT 11322

QY 357 IleTyAlaIleGlyLeuArgGluThrLysAspIleAsnAspAsnIleIlePheGln 376
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11323 GATATGATTAATAATACAGAGTAATTTAGGTATAGT---ATATTAGT 11379

QY 377 IleGlnProMetAsnAsnThrTyTyTyAlaSerGlnIlePheLysSerAsnPheAsn 396
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11380 ATAAAAATATGATGTCACAAATAAATATGTTAGTCAAAATTTCTCTAGTTTATG--- 11436

QY 397 GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyArgPheArgLeuGlyGly 415
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11437 ---AAAAACAATGCTTCTAGCAGATATATAAACTTGGAGATTTCTTTT--- 11487

QY 416 AspTptTyArgHisAsnTyLeuValProThrValLysGlnGlyAsnTyAlaSerLeu 435
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11488 -----GAAATGCATACACGCCA---GTTGCAGTAACATATATGACACAAA 11532

QY 436 LeuGluSerThrSerThrHisTrpGlyPheVal 446
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 11533 CTATTCAACTTCATCTTTTGGAAATTTATT 11565

RESULT 12
CBNTD
LOCUS          CBNTD          3952 bp      DNA      linear      BCT 17-MAY-1995
DEFINITION    Clostridium botulinum gene for neurotoxin type D.
ACCESSION    X54254
VERSION      X54254.1 GI:40395
KEYWORDS     neurotoxin; neurotoxin type D.
SOURCE       Clostridium botulinum.
ORGANISM     Clostridium botulinum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium
1 (bases 1 to 3952)
Niemann,H.
Direct Submission
Submitted (03-ADG-1990) Niemann H., Institut f Mikrobiologie,
Bundesforschungsanstalt f Viruskrankheiten, Postfach 11 49, D 7400
Tuebingen, F R G
REFERENCE 2 (bases 1 to 3952)
Binz,T., Kurazono,H., Popoff,M.R., Eklund,M.W., Sakaguchi,G.,
Kozaki,S., Kriedelstein,K., Henschen,A., Gill,D.M. and Niemann,H.
Nucleotide sequence of the gene encoding Clostridium botulinum
neurotoxin type D
Nucleic Acids Res. 18 (18), 5556 (1990)
JOURNAL      Nucleic Acids Res. 18 (18), 5556 (1990)
MEDLINE      91016853
PUBMED       2216736
REFERENCE 3 (bases 1 to 3952)
Okamoto,R., Macfarlane,J.L. and Wolstenholme,D.R.
Evidence for the frequent use of TTC as the translation initiation
codon of mitochondrial protein genes in the nematodes, Ascaris suum
and Caenorhabditis elegans
Nucleic Acids Res. 18 (20), 6113-6118 (1990)
JOURNAL      Nucleic Acids Res. 18 (20), 6113-6118 (1990)
MEDLINE      91045077
PUBMED       2235493
FEATURES
source       Location/Qualifiers
1..3952      /organism="Clostridium botulinum"
             /strain="BVD/-3"
             /db_xref="taxon:1491"
31..35      /note="ribosome binding site"
47..3877    /note="neurotoxin type D"
             /codon_start=1
             /transl_table=11
             /protein_id="CAA38175.1"
             /db_xref="GI:40396"
             /db_xref="SWISS-PROT:P19321"
             /translation="MTWPKDFNYSDFVNDNLDILYRIFQNKLIITPVKAFMITQNIW

```

BASE COUNT 1597 a 398 c 610 g 1347 t  
ORIGIN

# Alignment Scores:

Pred. No.: 9.24e-50 Length: 3952  
Score: 867.50 Matches: 195  
Percent Similarity: 61.64% Conservative: 83  
Best Local Similarity: 43.24% Mismatches: 146  
Query Match: 36.28% Indels: 27  
DB: 1 Gaps: 11

US-09-910-186A-10 (1-450) x CBNTD (1-3952)

QY 2 ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
Db 2561 ACAAGCCTTTTAAATATTTTTCATATACATAATCTTTTATAAAGATATAATAT 2620  
QY 22 GluTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
Db 2621 GAATATTTCAATAGTATTAATGATTCAAAATTTTGAGCTTACAAAACAAAAATGCT 2680  
QY 42 LeuValAspThrSerGlyTyThrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
Db 2681 TTAGGGATACATCAGATATAATGCAGAGTGGGTAGGAGATATGTTCAACTTAT 2740  
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
Db 2741 ACGATATATACAAATGACTTTAAATTAAGTAGTTCAGGAGAT-----AAAATTATA 2791  
QY 82 ValThrGlnAsnGluAsnIleValTyAsnSerMetTyThrGluSerPheSerIleSerPhe 101  
Db 2792 GTAAATTTAAATAATAATATTTTATATAGCGCTATTTATGAGAACCTCTAGTGTATTT 2851  
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyThrIleIleAspSer 120  
Db 2852 TGGATTAGATATCTTAAAGATTTAACTAATTCATAATGAATATAACAATAAACAAGT 2911  
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140  
Db 2912 ATAGAACAAATCTGGGTGAAATATGATTAGGAAATGCAATATAGATAGATGATTTA 2971  
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyThrAspIleSerAsnAla 160  
Db 2972 CAAGATGTTAATAGAAAGTATAAAGTTTAAATTTTTCATTATAGTGAATCAATAGTCAT 3031  
QY 161 ProGlyTyThr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179  
Db 3032 ACAGGATATACAAATAAATGCTTTTGTACTACTAATAATAATAATGCGGTATATG 3091  
QY 180 LysIleTyThrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199  
Db 180 LysIleTyThrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199

VIPERSSDTPNPSLSKPPPTSKYQSYDPSYLSLSTDEQKDFLKGIIKLFRIINERDI  
GKLNIILYVGGFFMGDSPTPEDTDFTRHTTNIAVEKFPENGSKWVTNIITPVSILFEG  
PLNDIYTAASLTQQQNPSEFGFGLSILKVAPEFLILFSDVTSNQSAVLGKSI  
FCMDPIYALMHETLSHLQLYGNIPIPSDKRIRPQVSEGGFSQDGNVQVELEYTFGGI  
DVPIIPIERSOLREKALGHYKDIKRLNINKITIPSSWISNDIKYKIFSEKYNFDK  
DNTGNEVNDKFNLSYDLTNMSEVYSSOYNVKNRTHYFSRHVLPVFNILDDNI  
YTRDGFNLTKNGFNIENSGONIESPALOKLSSESVVDLFTKVCRLTKNSRSDSTC  
IKVNNRPLPVADKDSIQEIEFENKIIDETNQNYSKFSLDESILGQVPIINPEIV  
DPLPNVMEPLNGEEIVFYDDITKYVDYLYLSYIIESQKLSNNENILITTSVEE  
ALGYSNKIYTFPLSLAEKRGVQAGFLNVAENEVDFTTNIMKKDLDKISDVSVI  
IPIYGALNIGNSALRGNFQAFATAGVAFLEGGPEFTIPALGVFTFYISQIEREKI  
IKYIENLQVRKRWKDSYQMWSNWLSRTTFQFNHINYQYDSLSYQADAKIKDL  
EYKYSKDKENIKSOVENKSLDVKISEAMNNINKIRECSVTYLFKNMLPKVIDE  
LNKFDLRTTELINLDSNIIILVGEVDLKKAKNESFENTMPTNFYSYTNLSLKD  
INIEFNSINDSKILSQNKALYPTSGYNAEVRVGDVQNLNTIYNDFKLSGDKI  
IIVLNNILYSIIENSVFSFWIKSKDLTNSHNEYTIINSIQNSQKGLIRQNGIE  
WILODNRKYSLIFDYSLSLSTGYTNKFFVTIINIMGYMKLYIINGELKSOKE  
WGLVDKDTIYFIDENIDENQMLWDFNIFSKELSDENIYVEQIILRNLIKDY  
WGNEPLFDTEYIINDNYIDRYAPESNVILVQVDPDRSKLYTGNPTIKSVSNKPY  
SRILGDNITLHMLYNSRKYMIROTDITYATQGGECSONCVYALKLQSLNGYIGI  
FSIKNYSKNKYCSQIFSSFRENTMLADIYKWPWFSPKNAVTPVAVTNYETKLLSTS  
SFWKFIISRDPGWVE"



```
Db 3092 AACTTTATATAAATGGAGCAATTAAGCAGAGTCMAAAATTTGAAGATTTAGATGAGGTT 3151
QY 200 AspPheSerLysThrIleThrPheGluLeuAsnLysIleProAspThrGlyLeuIleThr 219
Db 3152 AAGTTAGATAAAACCATAGTATTGGAATAGATAGATATAGAT----- 3196
QY 220 SerAspSerAspAsnIleAsnMetThrIleArgaspPheThrIlePheAlaLysGluLeu 239
Db 3197 -----GAGAAACAGATGCTTTGGATTAGATTTTAAATATTTTCTTAAAGNATTA 3247
QY 240 AspGlyAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAsp 259
Db 3248 AGTAATGAGATATTAAATTTATATGATGAGGACAAATATTAAAGAAATGTTATTAAAGAT 3307
QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
Db 3308 TATTTGGGAAATCCCTTTGAAGTTTGCATACAGAAATATTATATTATGATGATTAATATA 3367
QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsn 299
Db 3368 GATAGTATATAGCCACTGAAGTAATGCTACTGTGCTGTTCTAGTATCCAGATAGATCT 3427
QY 300 AspPheAsnGluGlyTyrLysIlelelelelelelelelelelelelelelelelelele 319
Db 3428 AAATTATATAGTGAATCCTTACTATTAAATCAGTATCTGTAAGAATCCCTTATAGT 3487
QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
Db 3488 AGAATTTTAAATGGAGATTAATAATCTTCATATGTTATATATAGTAGGAATATATG 3547
QY 340 LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHis-----SerThrGluAsp 356
Db 3548 ATAATAAGAGATAGTACTACATATATATGACACACAGGAGGAGAGTGTTTCAAAATGT 3607
QY 357 IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln 376
Db 3608 GTATATGCAATTAATAATACAGAGTAATTTAGTAATTTATGTTATAGT---ATATTAGT 3664
QY 377 IleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsn 396
Db 3665 ATAAAAATATTGTATCTAAAAATAATATTGTACTCAAAATTTTCT---TCTAGTTTATAGG 3721
QY 397 GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGly 415
Db 3722 ---GAAATACAAATGCTTCTAGCATATATATAAACCCTGGAGATTTCT----- 3769
QY 416 AspTrpTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435
Db 3770 -----TTTAAAAATGCATACAGCCAGTCGAGTAACCT-----AATTATGAACAAAA 3817
QY 436 LeuGluSerThrSerThrHisTrpGlyPheVal 446
Db 3818 CTATTATCACTTCATCTTTTGGAAATTTATT 3850
RESULT 13
S49407
LOCUS S49407 4087 bp DNA linear PHG 08-MAY-1993
DEFINITION type D neurotoxin [bacteriophage d-16 phi, host = C. botulinum,
S49407 type D, CB16, genomic, 4087 nt].
ACCESSION S49407
VERSION S49407.1 GI:260238
KEYWORDS Bacteriophage d-16 phi CB16 host = C. botulinum.
SOURCE Bacteriophage d-16 phi
ORGANISM Bacteriophage d-16 phi
VIRUSES:
1 (bases 1 to 4087)
Sunagawa,H., Ohshima,T., Watanabe,T. and Inoue,K.
AUTHORS The complete amino acid sequence of the Clostridium botulinum type
TITLE D neurotoxin, deduced by nucleotide sequence analysis of the
encoding phase d-16 phi genome
JOURNAL J. Vet. Med. Sci. 54 (5), 905-913 (1992)
MEDLINE 93042276
PUBMED 1420572
```

```
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gisbseq 118639] from the original journal article.
This sequence comes from Fig. 4.
FEATURES
Location/Qualifiers
1. 4087
/organism="Bacteriophage d-16 phi"
/db_xref="taxon:12409"
/note="type: D"
182..4012
/gene="type D neurotoxin"
182..4012
/gene="type D neurotoxin"
/note="Method: conceptual translation with partial peptide
sequencing. This sequence comes from Fig. 4"
/codon_start=1
/product="type D neurotoxin"
/protein_id="AAB24244.1"
/db_xref="GI:260239"
/translation="MTWPKVDNFYSDPVNDNDILYRIPQNKLTITPVKAFMITQNIW
VIPERFSDTNPISLSPKPRPTSKYQSYDPSYLSYSDDEQKDFLKGILKIFKRNERDI
GKLEINLVGSPMGDSSTPDEFDFRTTNTIAVEKPKENGSKVNTIITPVLFIFG
PLNILDYATSLATLOGQSNPSEGFCTLSILKVAPEFLTFTSDVTSNOSSAVLGKSI
FCMDPVIALMHLEIHLHQLYGNIIPSDKRIKRVQSEGFSDQSPNVQSEELYTFGGL
DVEIIPQERSQLREKALGHYKDIKRNININKTIPSSWISNIDKTKKIFSERINFDK
DNTGNFVNIDKFNSLYDLTNVMSVYVSYQYVNRNTHYFRHRLPVFANLDDNI
YTRDGFNLTKNGFIENSGNIERNPALQKLSSEVYDLTKVKLYLPVANSRDDSTC
IKVKNRNPYVADKDSIQEIFENKIITDNTVQNTSDNPSLSDILQGVIPNPIV
DPLPNVMEPLNLPGEIYFDITKYVDYLSYVLESOKLSNVENTITLTSVEE
ALGYSNKIYTFELSLAEKVKNGVQAGLFLANAEVVEEDFTNIMKKDLQKLSQSVKI
IPIGLNALIGNSALGNFKQAPATAGVAFLLGEGFEFTIPALGVTFVSSYQERKI
IKTENCLEQVKKWKDSYQMVNSWLSRITTFQNHINQYDLSLQADAKLIDL
ENKYSQSDENIKSQVENKLSLDVKSISAMNNINKFIRECSVTYLFKNMLPKVIDE
LNKPDLTKEILNIDSHNIIIVGEVDRLKAKVNFSEFNTPFNTITNNSLLKDI
IYEVNSNDKLSLQNKKNALVDTSYNAEVRVGDVQNTIYTNDFKLSGGDKI
IVNNNINLSAIYENSVSSEFWKISKDLTNSHNEYTIINSIEONSGWKLICRNGNIE
WILDQVNRKYSLLFEDYSLSHTGYTKWFEVITNNIMGYMKLVINGELKQSOXIE
DLQVKLDITVIGIDENIDENQMLIRDNIFSKELSNEDINIVVEGQLIRVINDY
WGNPLKDEYIINDYIAPESNVLVLPDRSKLYTGNTPIITKRSVSDKNPY
SRILGNDNIILHLYNRSKYIMIRDTTIYATGEGCSQNCVYALKLQSNLNGYIGI
FSIKNIKNKYCQISFSEFRNMTLLADIYKPFRESFKNAYTPVAVTNYETLLSTS
SEWKFISRDPGWE"
```

BASE COUNT 1646 a 413 c 633 g 1395 t

ORIGIN

Alignment Scores:

Pred. No.:	9-56e-50	Length:	4087
Score:	867.50	Matches:	195
Percent Similarity:	61.64%	Conservative:	83
Best Local Similarity:	43.24%	Mismatches:	146
Query Match:	36.28%	Indels:	27
DB:	7	Gaps:	11

US-09-910-186A-10 (1-450) x S49407 (1-4087)

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsn 21

Db 2696 ACAATCCCTTTTAAATATTTTTCATATACTAATAATCTTTTATAAAGATATAATAAT 2755

QY 22 GlutTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41

Db 2756 GRATATTTCAATAGTATTAATGATTTTCAAAATTTTGGAGCTTACAAAACAAAAAATGCT 2815

QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61

Db 2816 TTAGTGGATACATCAGGATATATATCGAGAGTGGAGGATAGGATATGTTCAACTTAAT 2875

QY 62 ProIlePheProPheAspPheLysLeuGlySerGlyGluAspArgGlyLysValIle 81

Db 2876 ACGATATACAAAATGACCTTAAATTAAGTAGTAGTCCAGAGAT-----AAAATTATA 2926

QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101

Db 2927 GTAAATTTAAATAATATATTTTATATAGCGCTATTTATGAGAACTCTAGTGTAGTTT 2966



```
gene
CDS
    WKLEYISSNGFRPESWAENKVLAYNDVGFIVLSSSSNNSIWNPIKTAINSYIICTL
    SINVTAWTIIDNNNNITDQPIILNLPNFDINNQQILKLEKL"
    complement(3030..3905)
    /gene="hai"
    complement(3030..3905)
    /gene="hai"
    /codon_start=1
    /transl_table=11
    /product="HA1"
    /protein_id="BAA75082.1"
    /db_xref="GI:4579736"
    /translation="MGNNGNSQTNANDLRNNEVEFFISPSNNINKVLDKISQSEVKLWN
    KLSGANKVKIYDITNKQAVKIKVMONTSLILTNAPLSSVSVKDTINGDNOYWLLO
    NYSRWYIRYNNPNLVLYQINDITLWYSTOTSSNQPFKFSNCITVILNRRNKLO
    TOLNSDFSLKNSOILVILWOFDSROKWLLEYNETKSAVTLKCOENRYLTIQN
    SNNVYEQSTDSLIQVWNLNLDNASKYILNLDQTNRYLDVINSQIANGTHRVVD
    SYHNTNQWMIINLI"
    4153..7743
    /gene="ntnh"
    4153..7743
    /gene="ntnh"
    /codon_start=1
    /transl_table=11
    /product="NTNH"
    /protein_id="BAA75083.1"
    /db_xref="GI:4579737"
    /translation="MDINDOLNINSPVDKNVIVVRARKNTFFKAFKVPNIWVAPE
    RYVEPIDIAEYKLDGIVDSNFDQSDSERENFLQAIILIKLRINNTISGKQLLSLI
    STAPFVYGVGGYSSPNITFTGKTPKSNKLSNLTSTIPPPGGYRETNVIFSON
    NKNYASNIVFPGSNIVSNVYIYKKDAENGMTMAEIVFPQLIYKYNKFIDP
    AMETKLIISLFIKIDPSDLVFPYRTELNDKQFSQNLIDILLISGVPDLEFI
    NLNPFYTNFYFNSIKMFYKNIYKTEIEGNAIGNDKLKKOKFOINQVDIWNK
    NLNFCQSPNIIIDPFSNALKHFYKQYITDNYINGVINGQINIKLPISKN
    TNLISKEPVNVLVNNENISLMSKNIVDGLKSTEDFYSTKIPYNEEYEFNDSO
    NPLNNISIEVDISPIIIDIINPYKONSDMLFTQITSMTEVTHTALSINYLQAOI
    TNNENFLSDQSKVSVSSDKSLVYSFLDNLMSYLETINKDGPIDTKVYVWLKEVF
    KNYSFIDLNDQSDMCGINEVLPWFGKALNLTNSNVEEYQDSGALSLSKKDLN
    REPNIIDDSGLDGLKANKLFEIYSKNIVFKIYFSLDQWMTETYSQFE
    LICKAKOSILASQSVKQIVONKFTDLSKASIPDCLKIRETTEFTIDLSNESOIS
    MNRVNDLNKASQCFVEDIYKPIYSIMEKYININIKTREFIQRTNIDNREKILI
    NSYTKIDPDKFDIQIKNFNSQVQVMKLEISPYQLLFASKSPNINIEDISGK
    NLTIQYTESIELVYNGVESLYLKSNETIKFSNKEFTNGLNANFTICFWLFTQKND
    DKTLLGNKNKNWEYTFEIDNGLVFEIIDSNGNESVYLSNIINDNYYIISVDRL
    KQDLILFNKNKNVAVSIDOLSVYSTNIISLVNKNNSLVVEELSVDNPITSEVIR
    NYFSKLSNSTRSSKSLLEYKNYQLVNYFPEISLYEVNDNNSKYLKNTDGINI
    SVYFKLINIDESGYVQKWEDECIICVDGTEKILDOIENNRKIQIYVSKDNKKKITY
    NTDLFRPDCITFYNDKIFSLSURDGDYNWICNDNNKVPKGAHLWILES"
    7758..11584
    /gene="ntx"
    7758..11584
    /gene="ntx"
    /codon_start=1
    /transl_table=11
    /product="NTX"
    /protein_id="BAA75084.1"
    /db_xref="GI:4579738"
    /translation="MTWPVKDFNYPDPNDNDILYRIPONKLTTPVKAFTINW
    VIPRFSDDNPSIKPRTSKVQSYVDPSYLSDEQDKTLGKIKLKRINERDI
    CKKILNVLVSGSPMGSSSTPDEFTDTRHTINIAVEKPENGSKWNTNITISYLFIF
    PLPNILDTASLTQQGQSPSPFEGFTLSILKVAPEFLITFSDVTSNQSQSAVLGSKI
    FCMQPDVALHETLSHQLYGINIPSKRIRIOPVSEGFQSDQPNVQEEELYTGGI
    DVEIIPQIERSQLREKALGHYKDIARLNNINKTIPSSWISNDIKYKIFSEKYNFDK
    DNTGNVYIDIKNSLYSDLTVMSEVYVSQYNKVRNTHYFSRHYLPVPFANILDDNI
    YTIKDFGNLTNKGNIENSGONTERNPALOKLSSESVDLFTKVCRLTKNRSRDSVC
    IKVKNRILPVVADKDSISQIFENKIITDETNQVYSDKFSLDESILDGQVNPINPIV
    DPLLPVNMPELNLPGIEIVYDDITKYVDYLSAYYLESQKLSNNVENITLTSVEE
    ALGSKNITLPSLSAEKYNKVGAGLFLANAEVEDEFTINMKKDTLDKSDVSVI
    IPIYGPALINSGALRNFQAFAGVAFLEGFPEFTIPALGVYFTYSSQIQRKBI
    KTKENCLEQYKRWKQSYQWVSNWLSRITTFONTINQYMSLSYQADAKAKTLD
    EYKYSQEDKENIKSQVENIKSLNDSKISEAMNNINKFIRECSVTYLFKNMPLKVIDE
    LNKDLPRTKTELINLDSHNIILVGEVDRLLKAVKNSFENMPTNFIYFNKLSGDKI
    INEYFNISNDKSLNLOKNALVDTSYNAEVRVGDNOLNTIYFNDFKLSGGDKI
    IVNLNNNTLYSAIYENSVSPWIKSKDLINSHNEVTIINSIQNSGKWLICIRNGNIE
```

```
BASE COUNT      4511 a 1227 c 1612 g 4234 t
ORIGIN
Alignment Scores:
Pred. No.:      2.68e-49      Length:      11584
Score:          867.50      Matches:      195
Percent Similarity: 61.64%      Conservative: 83
Best Local Similarity: 43.24%      Mismatches: 146
Query Match:     36.28%      Indels: 27
Db:              7          Gaps: 11

US-09-910-186A-10 (1-450) x AB012112 (1-11584)
QY  2 ThrilleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10272 ACAATGCCCTTTTAATATTTTTCATATACATAATAATCTTTTATAAAGATATAATTAAT 10331
QY  22 GluTyrPheAsnAsnIleAsnAspSerIlysIleLeuSerLeuGlnAsnArgLysAsnThr 41
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10332 GAATATTTTCAATAGTATTAATGATTCAAAATTTTGTAGCTTACAAAACAAAAAATGCT 10391
QY  42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10392 TTAGTGGATACATCAGGATATAATGCAGAGTGAGGGTAGAGATAATGTTCAACITTAAT 10451
QY  62 ProIlePheProPheAspPheGlyLeuGlySerSerGlyGluaspArgGlyLysValIle 81
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10452 ACGATATATACAAATGACTTTAAATTAAGTAGTTCAGAGAT-----AAAAATATA 10502
QY  82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10503 GFAATTTTAATAATAATAATTTTATATAGCGCTATTTATGAGACTCTAGTGTAGTTT 10562
QY  102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10563 TGGATTAAGATATCTAAAGATTTAACTAAATTTCTCATATGAATGAATAATAATAACAGT 10622
QY  121 ValLysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10623 ATAGAACAAATCTCGGTGGAAATTTATGATTAGGAATGCCATATAGATGGAATTTTA 10682
QY  141 LysGlnAsnGluAspSerSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10683 CAAGATGTTAATAGAAAGTATAAAAGTTTAATTTTGTATTATAGTGAATCAATAGTCAT 10742
QY  161 ProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10743 ACAGGATATACAAATAATAATGTTTGTACTACTACTATAATAATAATAGGGGTATATG 10802
QY  180 LysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10803 AAACTTTATATAAATGGAGAATTAAGCAGAGACTCAAAAATTTGAAGATTTTAGATGAGTT 10862
QY  200 AsnPheSerLysThrIlePheGluIleAsnLysIleProAspThrGlyLeuIleThr 219
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10863 AAGTTAGATAAACCACTAGTATTGGGAATAGATGAGATAATAGAT----- 10907
QY  220 SerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeu 239
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10908 -----GAGAATCAGATGCTTTGGATTAGACATTTAATAATTTTCTTAAAGAATTA 10958
QY  240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAsp 259
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10959 AGTAATAGAGATATTAAATTTGTATATGAGGACAAATATTAAAGAAATGTTATTAAGAT 11018
QY  260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

D	b		11019	TATTGGGGAATCCCTTTCAAGTGTGCACAGAATAATTAATATTAAATCATTAATATAAACAATTAATA	11078
Q	y		280	ASnArGrMetyrMetTyrrAlaaSnSerArgInIleValPheAsnThrArgArgAsnAsnAsn	299
D	b		11079	GATAGCATTATAGCACCGTCGAAAGTAGTGACTTGTCTGTTTCAAGTACCAGATAGACT	11138
Q	y		300	ASPpheasnGLUGlyTyrrLysIllellellesilleglyeargglyasnthrasnaspthr	319
D	b		11139	AAtTtATatActGGAAATcctAttTattTaataactagttacgtgafgaagaaccttatagt	11198
Q	y		320	ArgVAlarGgiYgyiAspilleLeutyrrPheaSpwethrIleAsnAsnLySalaatyraSn	339
D	b		11199	AgAAnTTTTAAATGGAGATAATAAATCTTCATATGTATATAANTAGTAGGAATAATATG	11258
Q	y		340	LepPhemEtyLasngluThrmetyrrAlaaaSPashis-----SerThrGlutasp	356
D	b		11259	ATAAFAAGAGATAcTGTATACAATATATCAACACAAAGGAGAGAGTGTCCAAAATTGT	11318
Q	y		357	IletyrrAlalleGliyleuarqgluglnthrLysAspileasnAspniletlepheln	376
D	b		11319	CtATATgcATTAAATtACAGAGTAATTTAGGTAAATTCGTATAGT---ATATTAGT	11375
Q	y		377	IleGlnPrOmEtAscAsnThrTyrrTyrrAlaserGlnIlePheLyserAsnpHeaSn	396
D	b		11376	ATAAAAAATATTGTATCTAAATAATATATGTAGTCAAATTTC---TCTAGTTTTAGG	11432
Q	y		397	GlyGUaSnIIeserGIyleCyseRie--GlyThrTyrrpheargueucglycly	415
D	b		11433	----GAARATACAAGCTTCTAGCAGATATATAAACCTTGGAGATTCT-----	11480
Q	y		416	ASptTpTyrrgrHisAntyrLeuvalProThrVailySlnglnGlyAsntYralaserLeu	435
D	b		11481	-----TUTAAAAATGCCATCACGCCAGTTCAGTAACT- ----AATTATGAACAAAA	11528
Q	y		436	LeuGUSeSrThrSerThrHIsTrpGlpPheVal	446
D	b		11529	cTATTATCAATCTTCATCTTTTIGGAAATTTATT	11561
R	E	S	ULT	RESULT 15	
F	O	N	LOCUS	AF3000465	
D	E	F	INITION	AF3000465	
A	C	E	SSION	complete cds.	
V	E	R	SION	AF3000465	
K	E	S	WORDS	AF3000465.1 GI:15982936	
S	O	R	OURCE	Clostridium botulinum.	
O	R	G	ANISM	Clostridium botulinum.	
R	E	F	ERENCE	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
A	U	T	HORS	1 (bases 1 to 3876)	
T	I	T	LE	Kirma.N., Ferreira,J.L. and Baumstark,B.R. Characterization of six type A strains of Clostridium botulinum that contain type B toxin gene sequences Unpublished	
J	O	U	RNAL	2 (bases 1 to 3876)	
R	E	F	ERENCE	Kirma.N., Ferreira,J.L. and Baumstark,B.R. Direct Submission	
A	U	T	HORS	Submitted (28-AUG-2000) Department of Biology, Georgia State University, P.O. Box 4010, Atlanta, GA 30302-4010, USA	
J	O	U	RNAL	Location/Qualifiers	
F	E	A	TURES	1..3876	
S	O	R	CE	/organism='Clostridium botulinum' /isolate='588' /db_xref='taxon:1491' /country='USA: Ohio' /note='isolated from stool sample in 1976 type: Ab'	
D	E	S	CDS	1..3876	
				/codon_start=1 /transl_table=11 /product='type B cryptic neurotoxin' /protein_id='FABL11498.1'	

/db\_xref="GI:15982937"

/translation="MPVTTNNFNYNDPIDNNILMPEPPFRGMGRYTKAFKTTDIRIK  
 IIPERTFYKPDFENKSGISNRDVCYDYPDLNNDKKNIFLOTMIKLFNRKSK  
 PGEKLEMLINGIPIYGRHVPFLFEFTNVKLIISNPGEVERKKGIPANLI  
 FGKPLVLENETIDIGIIONRRFSGFGIMQMKFYEVYVNNVQENKSGASIFNRR  
 GYSDPALUMHELIVLHGLYIKVNDLPVINEKFFFMOSTDAIOABELYFGGD  
 PSLISPTDKSIYDKVLQNGRIGVDRLNKVLVCLSDPNININIKYKFKDKYFVEDS  
 EGRYSIDVESFDKLYKSLMFGTEFNIAENYKIKTRASYSDSLPPVKIKLNDLNTY  
 TIEEGFNISDNKKEZIRGQNAIKNQAYEISKEHLAVYKIQMKSVKAPGICIDVD  
 NEDLFFIADKNSFSDLSKNERIAINTQNNYENDFISNELIDTDLISKLELPSEN  
 KUTSDENYVYVTKQQAIFKLTIDENITFOYLSYQTFPDLDIRDISLTSSFDLALLFS  
 NYSVSFFSMYD IKTANKVAEAGLWAGWKQVINDFVBANKSSTMDKADI SLIYPI  
 GLALNTYQENKFNENAFETIAGASITLLEFPELLIPVGAFLFLESIDNKKIKIETI  
 NSALTTRDEKWDIMGYLIVAOMLSTVNTQFVITKEGMYKALNYOAAQLEBIKYKYN  
 YSEKERSNIDPNDVNSKLNREGINQADININNECSVSYLMMKMIPLAVKLDLDE  
 NDTURLNLIINDENYLDIGSAEYKESKYDYIKTSPIDFLSYTNTNTLILIFNKY  
 NSDILNLIINLRYKNDLUGYGAKEVYDGVKLDNKQKFLTSSANSKIRVTON  
 ONIIFNSMFDLSYSPWIRIKPKYKNDGQNTYIHNEYTILNCKMNSKYSISIRGNMI  
 WTILINDKTSKVSFFYSIKEDISYINRWFETITNSNNAKIYINGKLESIDIRLD  
 IREVIANDEIIFPKLDGINDORTQIWNKYSFENFELSQNSIEETIKYISSEYKDFW  
 GNPLWYKMYFNAPNAGKSIYIKLKKDSVGEILTRKYNONSKIYINRDLYIGKFI  
 TRKKSQSINDIDVBEKDYIYLDFFNLNQRWYMYKYFKPEEKLFAPISDSDEF  
 YNIQIKVEYDQTSYSCQLLPKKDBESTDEIGLIGHRFYESGIVFKYKDFKPCISKW  
 YLKEVKKRYSKLGSCNWFQIFDPSGWTG"

BASE COUNT 1611 a 367 c 616 g 1282 t  
 ORIGIN

Alignment Scores:

Prod. No.:	2.17e-32	Length:	3876
Score:	610.00	Matches:	168
Percent Similarity:	50.81%	Conservative:	84
Best Local Similarity:	33.87%	Mismatches:	154
Query Match:	25.51%	Indels:	90
DB:	1	Gaps:	20

US-09-910-186A-10 (1-450) x AF300465 (1-3876)

QY	2	ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn	21
Db	2503	AGTATACCATTTGATCTTCAACGTACTACTAAATACAACTACTACTAAGAAATTTAAT	2562
QY	22	GlutTyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr	41
Db	2563	AAATATAATACGATATTTTAAATAATTATCTTAAATTAAGATATAGGGATAATAAG	2622
QY	42	LeuValAspThrSerGlyTyThrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn	61
Db	2623	TTAATAGATTTATCAGGATATGGCGCAAGGTAGAGGTATATGATGGGGTCGAAGCTTAAT	2682
QY	62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgglyLysValIle	81
Db	2683	GAT-----AAAATCAATTTAAATTAACCTAGTTCAGCA-----AATAGTAAGATTAGA	2730
QY	82	ValThrGlnAsnGluAsnIleValTyThrAsnSerMetTyGluSerPheSerIleSerPhe	101
Db	2731	GTGACTCAAAATCAGATATCATATTAATAGTATGTCTCTGATTTAGCGGTAGTTT	2790
QY	102	TyrIleArgIleAsnLysTyr-----ValSerAsnLeu-----ProGlyTyr	115
Db	2791	TGGATAAGAAATACCTAAATAAAGATGATGGGTACAAAAATATATTCATAATCAAT	2850
QY	116	ThrIleIleAspSerValLysAsnAsnSerGlyTyThrSerIleGlyIleIleSerAsnPhe	135
Db	2851	ACAATAAATAATCTGATGAAAAATAATCTGGATGGAAAAATATCATTAGGGGTAAATG	2910
QY	136	LeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyAsp	155
Db	2911	ATAAATATGGACTTTAATTGATATAAATGGAAAAAATCAATCAGTATTTTTCAGATATAGC	2970
QY	156	IleSerAsnAsnAlaProGlyTyr---AsnLysTyThrPhePheValThrValThrAsnAsn	174
Db	2971	ATAAAGAAGATATATCATCAGATATATAATAGATGTGTTTTTGTACATTTACTATAAT	3030

QY	175	MetKcEgLYasNMeTlysleTyrrleAsnGlyLysleuAAspThrIleLysVallys	194
Db	3031	---TCGGAACAGCCTAAAAATTTATATTAAATGTAAGCTAGAAATCACATATATAGATATTAGA	3087
QY	195	GlueUthrGlrIleAsnPheserLysThrIleThrPheGlueIleAsnLysIleProAsp	214
Db	3088	GATATAAGAGAGATTATCTTAATGATCAAAATATAITTAATAATTAGAT	3135
QY	215	ThrGlyLeuIleThrSerAspSerAspAsnIleAsn	230
Db	3136	-----GGTAATATAGATAGAACACACAGTTCATTGTGGATGAA	3171
QY	231	AspPheTyrrIlePheAlaLysGlueAspGlyLysAspIleAsnIleLysPheAsnSer	250
Db	3172	TATTTTCAGTATTTTATATACGGAATTAAGTCATCAATAATTAAGAAATATATAAAAT	3231
QY	251	LeuGlnTyThrAsnValValLysAspTyrrTrpGlyAsnAspLeuA-gTyrrAsnLysGlu	270
Db	3232	CAATCATATAGCGAAATTTAAAGAGATTTTGGGAAATTCCTTTAATGTACAATAAGAA	3291
QY	271	TyrrTyrrMetValAsnIleAspTyrrLeuAsnArgTyrrMetTyrr	286
Db	3292	TATTTATATGTTTAAATCGGGGAATAAAAATTCATATATTAACCTAAAGAAGATTCATCT	3351
QY	287	SerArgGlnIleValPheAsnThrArgAsnAsnAsn	299
Db	3352	GTAGTGGAATTTTAAACGTAGCAAAATATAATCAAAATATCCAAATATATAAAATTATAGA	3411
QY	300	AspPheAsnGluClyTyrrLysIleIleLysArgIleArgGlyAsnThr	316
Db	3412	GATTTATATATGGAGAAATTTATATAGAGA---AAGTCAAAATTCCTCAATCTATA	3468
QY	317	AsnAspThrArgValArgGlyGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLys	336
Db	3469	AATGATCATATAGTTTAAAGAAGAATATATATATCATAGATTTTTTAAATTAATCAAA	3528
QY	337	AlaTyrrAsnLeuPheMetLysAsnGluThrMetTyrrAlaAspAsnHisSerThrGluAsp	356
Db	3529	GAGTGGAGAGATATATAG	3546
QY	357	IleTyrrAlaIleGlyLeuArgGluGlnThrLys	372
Db	3547	---TATAAATTTTAAAGAAGAGGAGAAATAATGTTTTAGTCTCCTATAAGTGATCT	3603
QY	373	IleIlePhe-----GlnIleGlnProMetAsnAsnThrTyrrTyrrTyrrAlaSer	398
Db	3604	GATGAGTGTTCACAACTACTATACAAATAAAGGAATATGATGAACAGCCACCATATAGTTGT	3663
QY	389	Gln---IlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly	407
Db	3664	CAGTGTCTTTTAAAAA-----GATGAAGAAAGTACTGATGATAGGATGATGTTGGT	3717
QY	408	ThrTyrrArgPheArgLeuGlyGly	415
Db	3718	ATTTCATCGTTTCTACGAATCTGGAATGTGATTTAAAGAGTATAAGATTTATTTTGTATA	3777
QY	416	---AspTrpTyrrArgHisAsnTyrrLeuValProThrValLysGlnGlyAsnTyrrAlaSer	434
Db	3778	AGTAAATGGTAC-----TTAAAGAGAGTAAAGAGAAACCATATATAATCA	3822
QY	435	LeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu	450
Db	3823	AAATATG-----GGATCTAATGGCAGTTTATTCCTAAAGATGA	3861

Search completed: November 7, 2002, 18:38:07  
Job time : 2772 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 16:30:36 ; Search time 253 Seconds  
(without alignments)  
4005.531 Million cell updates/sec

Title: US-09-910-186a-10

Perfect score: 2391

Sequence: 1 MTIFNIFSTNNSLLKDII.....NYASLLESTHGWGFPVSE 450

Scoring table:

BLOSUM62	
Xgapop 10.0, Xgapext 0.5	
Ygapop 10.0, Ygapext 0.5	
Fgapop 6.0, Fgapext 7.0	
Delop 6.0, Delext 7.0	

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09910186/runat\_04112002\_111616\_20826/app\_query.fasta\_1.647  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -NACLEN=2000000000  
-USER=US09910186 -CGCN\_1.1.125 -runat\_04112002\_111616\_20826 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -NAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_101002.\*

1:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID32/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID32/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2391	100.0	1371	21	AA287214	DNA encoding synth
2	2391	100.0	1371	22	AAA54486	Botulinum toxin hea
3	2386	99.8	3950	20	AA25521	DNA coding for mod
4	2319	97.0	1502	19	AAV30588	Clostridium botuli
5	1589	66.5	1208	21	AA298630	C. botulinum type
6	1589	66.5	1440	21	AA298631	Type D BoNT non-to
7	868.5	36.3	1374	22	AAA54487	Botulinum toxin hea
8	803.5	33.6	1469	19	AAV30591	Clostridium botuli
9	589.5	24.7	1400	21	AA287215	DNA encoding synth
10	589.5	24.7	1400	22	AAA54488	Botulinum toxin hea
11	586.5	24.5	2532	21	AA287218	DNA encoding nativ
12	586.5	24.5	3891	17	AA29244	C. botulinum type
13	586.5	24.5	4835	21	AA264582	BoNT/A neurotoxin
14	583	24.4	4017	22	AA258862	C. botulinum BoNT/E
15	581	24.3	3876	19	AAV30579	Clostridium botuli
16	578.5	24.2	1472	19	AAV30585	Clostridium botuli
17	576	24.1	1463	19	AAV30584	Clostridium botuli
18	575.5	24.1	1347	21	AA287217	DNA encoding synth
19	575.5	24.1	1368	22	AA54491	Botulinum toxin hea
20	575.5	24.1	1535	19	AAV30596	Clostridium botuli
21	565	23.6	1526	19	AAV30580	Clostridium botuli
22	558.5	23.4	1402	17	AA29246	Type A neurotoxin
23	558.5	23.4	1402	19	AAV30572	Clostridium botuli
24	558.5	23.4	1546	19	AAV30575	Clostridium botuli
25	557.5	23.3	1317	21	AA287220	DNA encoding nativ
26	557.5	23.3	1330	19	AA29245	Type A neurotoxin
27	557.5	23.3	1330	19	AAV30571	Clostridium botuli
28	557.5	23.3	1332	22	AAA54482	Botulinum toxin hea
29	557.5	23.3	1338	21	AA287216	DNA encoding synth
30	557.5	23.3	1351	19	AA2870576	Clostridium botuli
31	556.5	23.3	1323	22	AAA54483	Botulinum toxin hea
32	556.5	23.3	1326	22	AAA54484	Botulinum toxin hea
33	553.5	23.0	1278	22	AA54489	Botulinum toxin hea
34	550	22.8	1547	19	AAV30581	Botulinum toxin hea
35	546	22.8	1293	18	AA148100	Immunogenic type F
36	546	22.8	1313	18	AA148101	Immunogenic type F
37	546	22.8	1314	22	AAA54499	Botulinum toxin hea
38	546	22.8	1317	22	AAA54490	Botulinum toxin hea
39	542	22.7	1341	21	AA287213	DNA encoding synth
40	542	22.7	1341	22	AA54485	Botulinum toxin hea
41	539.5	22.6	3509	19	AAV26289	Recombinant botuli
42	531	22.2	1317	21	AA287216	DNA encoding synth
43	511.5	21.4	1460	19	AAV30593	Clostridium botuli
44	492	20.6	1347	22	AAA54588	Sequence encoding
45	471	19.7	1332	22	AAA54588	Sequence encoding

# ALIGNMENTS

RESULT 1	
AA287214	
ID	AA287214 standard; DNA; 1371 BP.
XX	
XX	AA287214;
AC	
XX	
DT	08-MAY-2000 (first entry)
XX	
DE	DNA encoding synthetic BoNT serotype C (BoNTC) Hc fragment.
XX	
KW	Botulinum neurotoxin; heavy chain; BoNT; serotype C;
XX	C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW	VEE; botulinism; vaccine; diagnosis; drug screening; ds.
XX	
OS	Clostridium botulinum.
XX	Synthetic.
FH	Key Location/Qualifiers

```

FT CDS 10..1362
FT /*tag= a
FT /product= "Synthetic botulinum neurotoxin serotype C
FT (BoNTC) heavy chain C-terminal fragment (Hc)"
XX WO200002524-A2.
XX
XX 20-JAN-2000.
XX
XX 09-JUL-1999; 99WO-US15570.
XX
XX 10-JUL-1998; 98US-0092416.
XX 12-MAY-1999; 99US-0133870.
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh WT, Smith L;
XX WPI: 2000-160827/14.
XX P-PSDB; RAY77136.
XX
XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
XX toxin serotypes A-G, is used for inducing an immune response against
XX botulinum .
XX
XX Disclosure; Page 41-42; 54pp; English.
XX
XX The invention relates to novel vaccines that induce a protective immune
XX response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
XX and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
XX DNA construct comprising a vector, and at least one nucleic acid
XX fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
XX serotypes A-G. In preferred embodiments of the invention, the vector is
XX a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
XX this vector results in the production of large amounts of a protein
XX encoded by a sequence cloned into the replicon. The constructs are used
XX to produce vaccines against botulism. The proteins can also be used as
XX diagnostic tools for the diagnosis of botulism. The transformed host
XX cells can be used to analyse the effectiveness of drugs and agents which
XX inhibit toxin effects. The vaccine currently used against botulism is
XX dangerous and expensive to produce, and contains formalin, which is very
XX painful for the recipient. Also, the vaccine is incomplete, in that only
XX 5 of the 7 serotypes are represented in the formulation. The novel
XX vaccine of overcomes these problems, as it is easily purified, and
XX available in large quantities. It is also expressed in the lymph nodes
XX for a better immune response. Sequences AA287212-287217 represent
XX synthetic DNA sequences encoding BoNT Hc fragments used in the present
XX invention. These were optimised for codon usage for expression in yeast.
XX
XX Sequence 1371 BP; 397 A; 416 C; 265 G; 293 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1.42e-205 Length: 1371
XX Score: 2391.00 Matches: 450
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 21 Gaps: 0
XX
XX US-09-910-186a-10 (1-450) x AA287214 (1-1371)
XX
XX 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLeuAspIleIle 20
XX
XX 10 ATGACCATCCCATTCACATCTCTCTACACCAACAACCTCCCTGTGTGAAGGACATCATC 69
XX
XX 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
XX
XX 70 AACGAGTACTTCAACAACATCAACGACTCCCAAGATCTGCTCCCTGCGACGCCGTAGAAC 129
XX
XX 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
XX
XX 130 ACCTTGGTCGACACCTCCGGTTACACGCCCGAGGCTCCCGAGGAGGGGTGACGTCACGCTG 199

```

```

Qy 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyClnAspArgGlyLysVal 80
Dy 190 AACCAATCTCCCATTCGACTTCAAGCTGGGTTCCTCCGGTGAGGACAGAGTAAAGTTC 249
Qy 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
Dy 250 ATCGTCACCCAGACGAGAACATCGTCTACAACTCCATGTACGAGTCCCTCTCCATCTCC 309
Qy 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120
Dy 310 TTCGSGATCAGAATCAACAAGTGGTCTCCAACTTCCAGGTTACACCATCATCFGACTCC 369
Qy 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Dy 370 GTCAGAACAACATCCGGTGTGTCATCGGTATCATCTCCAACCTCTCTGCTCTTCCACCTG 429
Qy 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAla 160
Dy 430 AAGCAGAACGAGGACTCCGAGCAGTCCATCACTTCTCTACGACATCTCCAACAACGCT 489
Qy 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
Dy 490 CCGTGTACACAAGTGGTCTTCGTCACCGTCAACAACATGATGGTAAACATGAAG 549
Qy 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
Dy 550 ATCTACATCAACGGTAAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAAC 609
Qy 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Dy 610 TTCCTCCAGACCATCCTTCGAGATCAACAAGTCCAGACACCGGTCTGTACCTCC 669
Qy 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240
Dy 670 GACTCCGCAACATCAACATGTGATCGGTGACTTCTACATCTTCGCCAAGGAGTTGGAC 729
Qy 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyr 260
Dy 730 GGTAAAGACATCAACATCCCTGTTCAACTCTTCGAGTACCAACCGTCTGCAAGGACTAC 789
Qy 261 TrpGlyAsnAspLeuArgTyrAsnLysGlyTyrTyrMetValAsnIleAspTyrLeuAsn 280
Dy 790 TGGGTTAAGCAGCTCGAGATACACAGAGGATCTACTAGTGTCAACATCGACTACTTTGAAC 849
Qy 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAsp 300
Dy 850 AGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACACAGAGTACACACACGAC 909
Qy 301 PheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320
Dy 910 TTCACGAGGGTTACAAGATCATCATCAAGCGTATCAGAGGTAACACCAACGACACACAGA 969
Qy 321 ValArgGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
Dy 970 GTCAGAGGTGTGATCCTGCTACTTCGACATGACTATCAACAAGGCTCACACCTG 1029
Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Dy 1030 TTCATGAAGAAGACGACCATGTACGCCGCAACACCACTCCACCGGAGGACATCTACGCCATC 1089
Qy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380
Dy 1090 GGTCTGCTGTAGCAGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCAAATG 1149
Qy 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Dy 1150 AACACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACITCAACGGTGAGACATC 1209
Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHis 420
Dy 1210 TCCGATATCTTCCATCGGTACCTACAGATTCGCTGCGGTGCTGCTGCTGCTGCTGCTGCTG 1269
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440

```

```

Db 1270 AACTACTTGGTCCCAACGTCAAGCAGGTAACTACGCTCTTGGTGGAGTCCACTTCC 1329
|||||
QY 441 ThrHisTrpGlyPheValProValSerGlu 450
|||||
Db 1330 ACCCACTGGGGATTGCTCCCACTCCGAG 1359
RESULT 2
AAA54486
ID AAA54486 standard; DNA; 1371 BP.
XX AC
XX AAA54486;
XX 11-APR-2001 (first entry)
DE Botulinum toxin heavy chain C-terminal coding sequence (serotype C).
XX Botulinum toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.
XX Synthetic.
OS Clostridium botulinum.
XX
XX Key Location/Qualifiers
FH 10..1362
FT CDS /*tag= a
FT /product= H_C peptide fragment
XX
XX WO200067700-A2.
XX
XX 16-NOV-2000.
XX
XX 12-MAY-2000; 2000WO-US12890.
XX
XX 12-MAY-1999; 99US-0133865.
XX 12-MAY-1999; 99US-0133865.
XX 12-MAY-1999; 99US-0133867.
XX 12-MAY-1999; 99US-0133868.
XX 12-MAY-1999; 99US-0133869.
XX 12-MAY-1999; 99US-0133873.
XX 29-JUL-1999; 99US-0146192.
XX
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX P-PSDB; AAB04092.
XX
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
XX Claim 2; Fig 5a; 73pp; English.
XX
XX Botulin neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX

```

```

SQ Sequence 1371 BP; 397 A; 416 C; 265 G; 293 T; 0 other;
Alignment Scores:
Pred. No.: 1.42e-205 Length: 1371
Score: 2391.00 Matches: 450
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-09-910-186a-10 (1-450) x AAA54486 (1-1371)
QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleile 20
Db 10 ARGACCATCCCATTCACATCTTCTCTACCAACCACTCCCTGTTGAAGGACATCATC 69
QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
Db 70 AACGAGTAGTCTCAACAACATCAAGACTCCCAAGACTCTCTCCCTGCAGAACCGTAAGAAC 129
QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
Db 130 ACCTTGGTCGACACCTCCGGTTACAACGCCGAGGTCTCCGAGGAGGTGACGTCACGCTG 189
QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80
Db 190 AACCCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCGCGTGAGGACAGAGTAAGTTC 249
QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
Db 250 ATCGTCACCCAGACGAGAACATCGTCTACCACTCCCACTGACGAGTCTCTCCATCTCC 309
QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120
Db 310 TTCTGATCAGAATCAACAAGTGGTCTCCAACTTGCCAGGTTACACCATCATCGACTCC 369
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 370 GTCAAGAAACAACCTCCGTTGGTTCATCGGTATCATCTCCAACCTCTCTGGTCTTCA 429
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
Db 430 AAGCAGAACGAGGACTCCGACGAGTCATCAACTCTCTACGACATCTCCACAACGCT 489
QY 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
Db 490 CCTGGTTACAACAAGTGGTCTCTGTCACCGTCACCAACAACATGATGGTAAACATGA 549
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
Db 550 ATCTACATCAACGGTAAAGTATGATCGACACCATCAAGTCAAGGAGTTGACCGGTATCA 609
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 610 TTCTCCAAGACCATCACCTTCGAGATCAACAAGATCCAGACACCGGTCTGATCACCTCC 669
QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240
Db 670 GACTCCGACACATCAACATGTGGATCGTGACTCTACATCTTCGCCCAAGGAGTTGGAC 729
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyr 260
Db 730 GGTAAGGACATCAACATCTCTTCAACTCTCTGAGTACCAACAACGTCGTCAGGACTAC 789
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
Db 790 TGGGTACGACCTGAGATACAAAGGAGTACTACATGGTCAACATCGCTACTTGAAC 849
QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300
Db 850 AGATACATGTAGCCCAACTCCAGACAGATCTCTTCAACACCAAGACGCTAAACAACGAC 909
QY 301 PheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320

```



Db 910 TTCAACGAGGTTACAGATCATCATCAACCGTATCAGAGGTACACCAACGACACACAGA 969  
Qy 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340  
Db 970 CTCAGAGTGGTGCATCTCTGCTACTTCGACATGACTATCAACAACAAGCCCTACAACCTG 1029  
Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
Db 1030 TTCATGAGACGAGACCATGTACGCGGCAACCACTCCACCGAGGACATCTACGCCATC 1089  
Qy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleGlnPromet 380  
Db 1090 GGTCTGGGTGAGCAGACCAAGGACATCAAGCACAACATCATCTCCAGATCCAGGCAATG 1149  
Qy 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
Db 1150 AACACACTTACTACTAGCTTCCAGATCTTCAGTCCCACTTCACCGGTGAGAACATC 1209  
Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHis 420  
Db 1210 TCCGTATCTGTCCTCATCGGTACTACAGATTCGCTGGGTGGTGCATGGTACAGACAC 1269  
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440  
Db 1270 AACACTTGGTTCCACTGTCAAGCAGGTAACACGCTCCCTGCTGGAGTCCACTTCC 1329  
Qy 441 ThrIstrpGlyPheValProValSerGlu 450  
Db 1330 ACCCACTGGGATTCGTCCTCCAGTCTCCGAG 1359

RESULT 3

AAx25521

ID AAx25521 standard; DNA; 3950 BP.

AC AAx25521;

XX 02-AUG-1999 (first entry)

XX DNA coding for modified botulinum toxin rBoNT/C.

XX Botulinum toxin; botulism; rBoNT/C; vaccine; drug delivery;

XX mutant; ss.

XX Clostridium botulinum.

XX Synthetic.

XX Key Location/Qualifiers

FT CDS 32..3907

FT /\*tag= a

XX WO9920306-A1.

XX 29-APR-1999.

XX PF 16-OCT-1998; 98WO-US21897.

XX PR 20-OCT-1997; 97US-0954302.

XX PA (UJJE-) UNIV JEFFERSON THOMAS.

XX Kiyatkin N, Maksymowych A, Simpson L;

XX WPI; 1999-302646/25.

XX DR P-PSDB; AAY05814.

XX Modified toxin useful for systemic delivery of oral vaccines and

XX therapeutic agents

XX Example 1; Page 26-31; 37pp; English.

XX This DNA sequence codes for a modified serotype C botulinum toxin,

XX termed rBoNT/C (see AAY05814), in which amino acids His-229, Glu-230

CC and His-233 of the native sequence are substituted by Gly, Thr and  
CC Asn, respectively, i.e. the zinc binding motif (see AAY05814) of  
CC the light chain holotoxin is modified, resulting in loss of  
CC endoprotease activity. DNA coding for the modified botulinum toxin  
CC was assembled from 3 separate toxin fragments using PCR and  
CC site-directed mutagenesis. The modified recombinant botulinum  
CC toxin maintains its ability to translocate from the gut into the  
CC general circulation but is non-toxic. It can be used as an oral  
CC vaccine for antigenic peptides including botulinum toxin (i.e. an  
CC oral vaccine for botulism) or for the oral delivery of other  
CC therapeutic agents to the general circulation.

XX SQ Sequence 3950 BP; 1589 A; 410 C; 616 G; 1335 T; 0 other;

Alignment Scores:

Pred. No.: 1 43e-204 Length: 3950

Score: 2386.00 Matches: 449

Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 99.79% Indels: 0

DB: 20 Gaps: 0

US-09-910-186a-10 (1-450) x AAx25521 (1-3950)

Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn 21

Db 2558 ACAATACCCCTTAAATATTTTTCATATATAATCTTTTATAAAGATATAATTAAT 2617

Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41

Db 2618 GAATATTTCATTAATATTAATGATTCAAAATTTTCAGCTACAAAACAGAAAAATACT 2677

Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61

Db 2678 TTAGTGATACATCAGGATATAATGCAGAGTGCAGTGAAGAAGGCGATGTCAGCTTAAT 2737

Qy 62 ProfilePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81

Db 2738 CCAATATTTCCATTTGCTTTAAATAGGTAGTTCAGGAGGAGTAGAGTAAGTTATA 2797

Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101

Db 2798 GTAACCCGAATGAAAATATTTGTATATAATTCATGTATGAAAAGTTTATGACATTAGTTT 2857

Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121

Db 2858 TCGATTAGAATAAATAATGGTAAGTAATTTACCTGGATATACATAATGATAGTGT 2917

Qy 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141

Db 2918 AAAATAACTCAGGTGAGTATAGGTATTTAGTAATTTTTCATATTACTTTAAAA 2977

Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161

Db 2978 CAAAATGAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAAAATATGCTCT 3037

Qy 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181

Db 3038 GGATACATAAATGGTTTTTTTGTAACTGTTACTAACAAATATGATGGAAATATGAAGATT 3097

Qy 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201

Db 3098 TATATAATGGAAAATTAATAGATACTATAAAAAGTTTAAAGAACTAACTGGAATTAATTTT 3157

Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221

Db 3158 AGCAAACTATAACATTTGAATAAATAAATTCAGATACCCGTTTGATTACTTCAGAT 3217

Qy 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241

Db 3218 TCTGATAACATCAATATGCGATAAGAGATTTTATATATATTTGCTTAAGAATTAGATGT 3277

Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261

```

Db 3278 AAAGATATTAAATATATTATTAATAGCTGCAATATATACTAATGTTGTAAGAGATTATTTGG 3337
QY 262 GIVASnSPLeuArgTyrAsnLysGluTyrMetValAsnLleAspTyrLeuAsnArg 281
Db 3338 GGAATGATTTAAGATATAAATAAGAAATATTATATGGTTAATAGATTATTTAAATAGA 3397
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPhe 301
Db 3398 TATATGATGCGAACTCAGCAAAATTCGTTTTTAATACACGTAGAAATATAATAGACTTC 3457
QY 302 AsnGluGlyTyrLysLleLleLysArgLleArgGlyAsnThrAsnAspThrArgVal 321
Db 3458 AATGAGGATATAAATATATAAATAAGAAATCAGAGGAAATCAAAATGATAGTAGTA 3517
QY 322 ArgGlyGlyAspLleLeuTyrPheAspMetThrLleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3518 CGAGGAGGAGATATTATATTTGATATGACATTAATAACAAGCATATAATTTGTTT 3577
QY 342 MetLysAsnGluThrMetTyrAlaAsnAsnHisSerThrGluAspLleTyrAlaLleGly 361
Db 3578 ATGAAGAATGAACATATGTCAGATATATCATAGTACTGAAGATATATATGCTATAGGT 3637
QY 362 LeuArgGluGlnThrLysAspLleAsnAspAsnLleLlePheGlnIleGlnProMetAsn 381
Db 3638 TTAAGAGAACAAACAAGGATATAATCATATATATATTTCAATACACCAATGAAT 3697
QY 382 AsnThrTyrTyrTyrAlaSerGlnLlePheLysSerAsnPheAsnGlyGluAsnLleSer 401
Db 3698 AATACTATATATATGCGATCTCAATATATTAATCAAAATTTTAATGGAGAAATATTCT 3757
QY 402 GlyLleCysSerLleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsn 421
Db 3758 GGAATATGTTCAATAGGTACTATCGTTTGTAGACTTGGAGGTGATTTGTATAGACAAAT 3817
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 3818 TATTTGGTGGCTACTGTGAAGCAAGAAATATGCTTCATTATTAGAATCAACATCAACT 3877
QY 442 HisTrpGlyPheValProValSerGlu 450
Db 3878 CATTTGGGGTTTGTACCTGTAAGTAA 3904
RESULT 4
AAV30588
ID AAV30588 standard; DNA; 1502 BP.
XX AAV30588;
AC AAV30588;
DT 07-DEC-1998 (first entry)
XX Clostridium botulinum type C1 toxin C fragment gene in pETHisb.
XX Antitoxin; vaccine; neurotoxin; toxin C; intoxication; immunogen;
KW botulism; Botc; ds.
XX Clostridium botulinum serotype C1 Stockholm strain.
OS Synthetic.
XX Key
FH Location/Qualifiers
FT CDS 108..1463
FT /*tag= a
FT FT
FT FT
PN WO9808540-A1.
XX
XX 05-MAR-1998.
XX
XX 28-AUG-1997; 97WO-US15394.
XX
XX 28-AUG-1996; 96US-0704159.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
XX
XX

```

```

PI Thalley BS, Williams JA;
XX
DR WPI: 1998-230234/20.
DR P-PSDB; AAW68397.
XX
PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX
PS Example 46; Page 339-341; 428pp; English.
XX
CC This is the DNA sequence of the Clostridium botulinum serotype C1
CC (Stockholm strain) neurotoxin fragment C gene contained in plasmid
CC pETHisb. The encoded BotC fragment C polypeptide (see AAW68397) has a
CC His-tagged N-terminal extension. The vector was used to express
CC native (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC cells. The invention relates to recombinant proteins derived from
CC C. botulinum toxins, especially type B and type E toxins. Methods
CC are provided which allow for the isolation of soluble recombinant
CC proteins free of significant endotoxin contamination. Preferred
CC hosts for production of recombinant proteins are E. coli, insect
CC cells and yeast cells. The recombinant toxins are used as
CC immunogens for the production of vaccines and antitoxins that are
CC useful in the treatment of humans and animals at risk of
CC intoxication with clostridial toxin.
XX
SQ Sequence 1502 BP; 600 A; 155 C; 249 G; 498 T; 0 other;
Alignment Scores:
Pred. No.: 4.67e-199 Length: 1502
Score: 2319.00 Matches: 439
Percent Similarity: 98.66% Conservative: 3
Best Local Similarity: 97.99% Mismatches: 5
Query Match: 96.99% Indels: 1
Dbs: 19 Gaps: 0
US-09-910-186A-10 (1-450) x AAV30588 (1-1502)
QY 4 ProPheAsnLlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspLleLeuAsnGluTyr 23
Db 152 CCATATCGAAGTCGTCATATGCGCTAGCTAGCTTTTAAAGATATATAATTAATGATA 211
QY 23 rPheAsnAsnLleAsnAspSerLysLleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43
Db 212 TTTCAATATATTAATGATTCAAAAATTTTGAGCCTCAAAACACAGAAAAATACTTTAGT 271
QY 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIl 63
Db 272 GGATACATCAGATATATAATGCAGAAGTGAGTGAGAGAGGCGATGTTGAGTTAATCCAAT 331
QY 63 ePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValLleValTh 83
Db 332 ATTTCATTTGACTTTAAATTAGGTAGTTTCAGGGGAGGATAGAGGTAAAGTTATAGTAAC 391
QY 83 rGlnAsnGluAsnLleValTyrAsnSerMetTyrGluSerPheSerLleSerPheTyrPil 103
Db 392 CCAGAATGAANAATATGATATATATCTATGATGATCAAAAGTTTACGATTTAGTTTGCAT 451
QY 103 eArgLleAsnLysTrpValSerAsnLeuProGlyTyrThrLleLleAspSerValLysAs 123
Db 452 TAGAATAAATAAATGGTAAAGTAATTTACCTGGATATATACTATAATTCATAGTTAAAAA 511
QY 123 nAsnSerGlyTrpSerLleGlyLleLleSerAsnPheLeuValPheThrLeuLysGlnAs 143
Db 512 TAACTCAGGTTGGAGTATAGGTATATTAGTAATTTTGTAGTATTACTTTAAACAAAAA 571
QY 143 nGluAspSerGluGlnSerLleAsnPheSerTyrAspLleSerAsnAsnAlaProGlyTy 163
Db 572 TCAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAAAATATGCTCTCGATA 631
QY 163 rAsnLysTrpPhePheValThrValThrAsnAsnMetMetClyAsnMetLysLleTyrIl 183
Db 632 CAATAAATGGTTTTTTTGTAACTGTACTAACATATGATGGGAAATATGAAGATTATAT 691

```

```

QY 183 eAsnGlyLysLeuLeuAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLy 203
DB 692 AAATGGAATAATATAGACTATATAAAAGTTAAAGAACTAACTGGAATTAATTTTAGCAA 751
QY 203 sPhrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223
DB 752 AACTATAACATTGGAATAAATAAATCCAGATACCAGGTTGATTACTTCAGATTCTGA 811
QY 223 pAsnIleAsnMetTrpIleArgAspPheTrpIlePheAlaLysGluLeuAspGlyLysAs 243
DB 812 TAACATCAATATGTGGATAACAGATTTTATATATTTGCTAAAGAAATTAGATGTAAGA 871
QY 243 pIleAsnIleLeuPheAsnSerLeuGlnIleThrAsnValValLysAspTrpIleGlyAs 263
DB 872 TATTAATATATATTTAATAGCTTGCAATATCTAATACTGTTGTTAAAGATTTATGGGAA 931
QY 263 nAspLeuArgTrpAsnLysGluIleTrpMetValAsnIleAspTrpIleAsnArgTrpMe 283
DB 932 TGATTTAAGATATAATAAAGAAATATATATGTTAATATAGATTTATTAATAGATATAT 991
QY 283 tTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnG 303
DB 992 GATCGCAACTCAGACAAATGTTTAAATACACGTAGAAATTAATAGACTTCAATGA 1051
QY 303 uGlyTrpIleLysIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgG 323
DB 1052 AGCATATAATATATAAAGAAATCAGAGAAATACAAATGATACAGTACAGAGG 1111
QY 323 yGlyAspIleLeuTrpPheAspMetThrIleAsnAsnLysAlaTrpAsnLeuPheMetLy 343
DB 1112 AGGAGATATTTATATTTTGATGACAAATTAATAACAAAGCATATAATTTGTTATGAA 1171
QY 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363
DB 1172 GAATGAACATGATGATGAGATATATCATGATACGACAGATATATATGCTATAGTTTAA 1231
QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleClnProMetAsnAsnTh 383
DB 1232 AGAACAAACAAAGGATATAAATGATATATATATATATTTCAATACAAACCAATGAATAAC 1291
QY 383 rTyrTrpTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyI 403
DB 1292 TTATATTATCCAGCTCCAAATATTAATCAATTTTAATGGAGAAATATTTCTGGAA 1351
QY 403 eCysSerIleGlyThrTrpArgPheArgLeuGlyGlyAspTrpTrpArgHisAsnTrpLe 423
DB 1352 ATGTTCAATAGTACTTATCGTTTTAGACTTGGAGGTGATTTGGTATAGACACAAATATT 1411
QY 423 uValProThrValLysGlnGlyAsnTrpAlaSerLeuLeuGluSerThrSerThrHisTr 443
DB 1412 GTGGCTACTGTGAAGCAAGGAATATGCTTCATTATTAATCAACATCAACTCATTTG 1471
QY 443 pGlyPheValProValSerGlu 450
DB 1472 GGGTTTGTACCTGTAGTGAA 1493

RESULT 5
ID AA298630
XX AA298630 standard; DNA; 1208 BP.
AC
XX
XX AA298630;
DT
XX 20-JUN-2000 (first entry)
DE C. botulinum type D toxin nucleotide sequence.
XX
KW Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;
KN protect; ds.
XX
CS Clostridium botulinum.
XX
PN WO200005252-A1.

```

```

XX 03-FEB-2000.
XX 20-JUL-1999; 99WO-IB01301.
XX 22-JUL-1998; 98ZA-0006538.
XX (AGRI-) AGRIC RES COUNCIL.
XX De Bruyn EE, Botha AD;
PI WPI: 2000-205375/18.
DR P-PSDB; AAY78982.
XX
XX Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin
XX type D, useful in vaccines for protection against botulism, comprises
XX at least one amino acid mutation not present in the wild type D
XX neurotoxins
XX
XX Claim 8; Page 54-57; 66pp; English.
XX
XX This sequence represents a nucleotide sequence of a synthetic gene
XX encoding a non-toxic immunogenic derivative of Clostridium botulinum type
XX D toxin (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle
XX and sheep, and usually results in the death of the affected or poisoned
XX animal. The non-toxic immunogenic fragments of the C. botulinum
XX neurotoxin are useful in vaccines to protect animals (e.g. humans,
XX cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic
XX fragments can be produced relatively simply and inexpensively
XX (specifically by fermentation techniques). As the fragments are not
XX toxic the risk to production staff is reduced.
XX
XX Sequence 1208 BP; 416 A; 203 C; 216 G; 373 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1,43e-133 Length: 1208
XX Score: 1589.00 Matches: 307
XX Percent Similarity: 84.16% Conservative: 33
XX Best Local Similarity: 75.99% Mismatches: 56
XX Query Match: 66.46% Indels: 8
XX DB: 21 Gaps: 4
XX
XX US-09-910-186A-10 (1-450) x AA298630 (1-1208)
XX
XX QY 50 AlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLys 69
XX DB 4 GCAGAAGTAGAGTTGAGGGAAACGTCAGTTGAATCTTCCATTTCATTGACTTTAG 63
XX QY 70 LeuGlySerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleVal 89
XX DB 64 TTGGTAGCTCTGCTGATGACCGTGGCAAAATATTGTAACCTCAAAACGAAATATTGT 123
XX QY 90 TyrAsnSerMetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpVal 109
XX DB 124 TATAATGCTATGATGATCTTCTCTATTAGCTTCTGGATCAGAAATATAATAAGGTT 183
XX QY 110 SerAsnLeuProGlyTyrThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIle 129
XX DB 184 TCAATTTTACCAGGTTATACCATCATCGACAGCGTTAAATAATTATTCAGGATGTTAT 243
XX QY 130 GlyIleIleSerAsnPheLeuValPheThrIleuLysGlnAsnGluAspSerGluGlnSer 149
XX DB 244 GGCATAATTAGCAATTTTCTAGTTTTCACACTAAAGCAAAACGAAATACGAGCAGGAC 303
XX QY 150 IleAsnPheSerTrpAspIleSerAsnAsnAlaProGlyTyrAsnLysTrpPhePheVal 169
XX DB 304 ATTAATTTTCTTAGACATTTTCAAGAACGCTGCTGGTTATTAATAATGGTCTTTGTA 363
XX QY 170 ThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAsp 189
XX DB 364 ACCATTACCAACACATGATGGGTAATATGATGATCTACATTACGGAAGAAAGCTAATCGAC 423
XX QY 190 ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle 209

```

Db	424	ACCATCAAGTGAAGAAATGACCGGTATCAACTTCTCAAGACCATAGCTTCAGATG	493
QY	210	AsnLysIleProAspThrGlyLeulleThrSerAspSerAspAsnIleAsnMetTrpIle	229
Db	484	ATAAAATCCCAACACAGGATTAATAACGCTGCTATTCTGACAAACATCAATATGTGGATT	543
QY	230	ArgAspPheTyrIlePheAlaLysIleuLeuAspGlyLysAspIleAsnIleLeuPheAsn	249
Db	544	AGAGATTTCATATTTTCGCAAAAGAAATTAGATGACAGGACGATTAATATCCTTTCAAT	603
QY	250	SerLeuGlnTyrThrAsnValIleValIleValIleValIleValIleValIleValIle	269
Db	604	TCATTTCATACACATAATGTTGTTAAGGATTAATGGGTAATGATTAAAGATATGATAAG	663
QY	270	GLUTYTYRMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGln	289
Db	664	GAGTACTACATGATTAATGTTAAATTACATGAATCGTTACATGTCTCTAAAAGGGTACGGT	723
QY	290	IleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIle	309
Db	724	ATCGTCTTCAACACACGTTAAATAACACAGGATTCACAGGAAGGGTACAAAATATCATATA	783
QY	310	LysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe	329
Db	784	AAAAGAATCCGTGGAAATACAAATGATACTAGAGTTCGTGCTGAGACGCTCATATTTT	843
QY	330	AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla	349
Db	844	ATAACACCATAGATAATAAGCAGTACTCTTTA-----GGCATGTATATAA	888
QY	350	AspAsnHisSerThr---GluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAsp	368
Db	889	CCGTCAAGAAACTTAGGGACACACCTTGTTCATTGGGTGCATTAGATCAACCGATGAC	948
QY	369	---IleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAla	387
Db	949	GAAATTCGTAATATGGATCATCTATTATCAACCATGATACGTTTCGATTACTACGCG	1008
QY	388	SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly	407
Db	1009	AGCCAATTATTTCTTCTAGCAACGCCAGCACCAATCGTTTAGGAATCATATCAATCGC	1068
QY	408	ThrTyrArgPheArgLeuGlyAsp--TrpTyrArgHisAsnTyrLeuValProThr	426
Db	1069	TCATATCTTTAACTGGGTGATGATTACTGGTTAACCAATGAATCTTGATCCCACTA	1128
QY	427	ValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheVal	446
Db	1129	ATTAATAATCGAACATACGCTTCATTACTGGAATCAACTTCTACACATGGGTCTTTGTC	1188
QY	447	ProValSerGlu	450
Db	1189	CCAGCGCTCTGAG	1200
RESULT	6		
AAZ98631			
ID	AAZ98631	standard; DNA; 1440 bp.	
XX	AAZ98631;		
AC			
XX			
DT	20-JUN-2000	(first entry)	
XX			
DE	Type D BoNT non-toxic immunogenic derivative nucleotide sequence.		
XX			
KW	Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison		
KW	protect; ds.		
XX			
OS	Clostridium botulinum.		
XX			
PN	W0200005252-A1.		
XX			
PD	03-FEB-2000.		

[illegible]

Db	478	ACATCAAAAGTGAAGAANTTGACCGGTATCAACTTCTCAAAGACCAATTACGTTTCAGATG	537
QY	210	AsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIle	229
Db	538	AATAAAATCCCGAACACAGGATTAAACAGTCTGATTCTGACAACATCAATATGTGGATT	597
QY	230	ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn	249
Db	598	AGAGATTCTATTTTCGAAAGAAGTATGATGACAAGGACATTAATATCCCTTCAAT	657
QY	250	SerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArygTyrAsnLys	269
Db	658	TCACTTCAGTACACATAATGTTGTTAAGGATTATGGGTAATGATTAAAGATGATAAG	717
QY	270	GluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGln	289
Db	718	GAGTACTACATGATTAAATGTAATTCATGAATGTTTCATGTTCTAAAAGGTACCGT	777
QY	290	IleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIle	309
Db	778	ATCGTCTTCAACACACTCGTAAAAATAACACAGCATTTTCAACGAAGGGTACAAAATTATCAT	837
QY	310	LysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe	329
Db	838	AAAAGAATCCGTGGAAATACAAATGATCTAGAGTCTCGTGTGAGACGTCGTATATTT	897
QY	330	AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla	349
Db	898	AATPACACAGCATAGATAAATAAGCAGTACTCTTTA-----GGCATGTATATAA	942
QY	350	AspAsnHisSerThr---GluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAsp	368
Db	943	CGTCGAAGAACTTAGGCACACACCTTGTCCATGGTGGCTAGTACACCGATGGAC	1002
QY	369	---IleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAla	387
Db	1003	GAAATTCGTAAATATGGATCATTCATTATCAACCATGTAATACGTTTCGATTACACGG	1062
QY	388	SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly	407
Db	1063	AGCCAATTATTTCTTTAGCAACGCCACGACCACCTGTTTAGGAATACATCAATCGGC	1122
QY	408	ThrTyrArgPheArgLeuGlyGlyAsp---TrpTyrArgHisAsnTyrLeuValProThr	426
Db	1123	TCATATTCCTTTAAACTGGGTGATGATTACTGGTTTAAACCATGAATACTTGATCCCACTA	1182
QY	427	ValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheVal	446
Db	1183	ATTAAATCGAACATTACGCTTCATTACTGGAATCACTTCTCACATTTGGTCTTTGTC	1242
QY	447	ProValSerGlu	450
Db	1243	CCAGCGTCTGAG	1254
RESULT 7	AAA54487		
ID	AAA54487		
AC	AAA54487		
XX	AAA54487		
XX	11-APR-2001 (first entry)		
DT	Botulism toxin heavy chain C-terminal coding sequence (serotype D).		
DE	Botulism; toxin; neurotoxin; heavy chain; recombinant expression;		
KW	recombinant vector; antigen; immune response; vaccine; bacterium;		
KW	infection; ds.		
XX	Synthetic.		
OS	Clostridium botulinum.		
XX	Key		
XX	Location/Qualifiers		
FT	10..1365		
CD	10..1365		

FT		/tag= a	
FT		/product= H_C peptide fragment	
XX			
PX			
NN			
PN			
PD		WC200067700-A2.	
PP			
PD		16-NOV-2000.	
PF			
PR		12-MAY-2000; 2000WO-US12890.	
PR			
PR		12-MAY-1999; 99US-0133865.	
PR		12-MAY-1999; 99US-0133866.	
PR		12-MAY-1999; 99US-0133867.	
PR		12-MAY-1999; 99US-0133868.	
PR		12-MAY-1999; 99US-0133869.	
PR		12-MAY-1999; 99US-0133873.	
PR		29-JUL-1999; 99US-0146192.	
XX		(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.	
PA			
XX			
PI		Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;	
PI			
XX			
DR		WPI; 2001-016048/O2.	
DR		P-PSDB; AAB04093.	
XX			
PT		New nucleic acids encoding the carboxy- or amino-terminal portions of the heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine against botulism	
XX			
XX		Claim 2; Fig 6a; 73pp; English.	
XX			
CC		Botulin neurotoxins are translated as a single 150 kDa polypeptide chain and then posttranslationally nicked, forming a dichain consisting of a 100 kDa heavy chain and a 50 kDa light chain which remain linked by a disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT) can be used in recombinant expression vectors and expressed in transformed cells to produce peptide antigens useful for eliciting an immune response to give protective immunity against botulinum neurotoxin, which causes botulism. The nucleic acids are expressible in a recombinant organisms such as Escherichia coli or Pichia pastoris. The use of recombinant nucleic acids are advantageous since it eliminates the need to culture large quantities of hazardous toxin-producing bacterium. Production yield from the genetically engineered product is also high and cost of production is lower. The nucleic acids can be derived from Clostridium botulinum serotypes A-G.	
XX			
SO		Sequence 1374 BP; 367 A; 423 C; 274 G; 310 T; 0 other;	
Alignment Scores:			
Pred. No.:	9.38e-69	Length:	1374
Score:	868.50	Matches:	197
Percent Similarity:	62.06%	Conservative:	86
Best Local Similarity:	43.20%	Mismatches:	145
Query Match:	36.32%	Indels:	28
DB:	22	Gaps:	11
US-09-910-186A-10 (1-450); X AA54487 (1-1374)			
QY	2 ThRIleProPheAsnIIEpHeSerTYrThRsnAsnSerLeuLeuLYsAspIIeIAsn 21		
Dd	49 ACATGCCATTCAACAVTTCTCTACACCACAACTCCTTGTTGAAGCATCATTCAC 108		
QY	22 GLUTYPHeASnaSnIieASnaSPserSYsIleLeuSeRLeUGInASnaRGlySAsnthr 41		
Dd	109 GAGTACTTCAACTCCATCAACGACTCCAGAATCTGTCTCGTAGACAACAGAACGC 161		
QY	42 LeuValASPThrSERGLYTyrASNAlaGlUvalSERGLUGluGYaspValGINLeuASN 61		
Dd	169 TTGGTCGACACCTCCGGRTTACACCCGAGTCAGAGTCGGTGACACGCTCCAGTTGAAC 228		
QY	62 PRoIIepHroPheASPpHeLYsLeuGLYSerSerSGlyGIuAsPARgGLYLysValIl 81		

```
Db 229 ACCATCTACACCAACGACTTCAAGTTGCTCTCCGGTGAC-----AGATCATC 279
QY 82 ValThrGlnAsnGluAsnValThrAsnSerMetTyrGluSerPheSerPhe 101
Db 280 GTCAACTTGAACAACAACATCTGTACTCCGCCATCTACGAACTCCTGTCTCTCT 339
QY 102 TpileArgGileAsnLysTrpValSerAsnLeuProGly---TyrThrIleLeuAspSer 120
Db 340 TGGATCAGATCTCCAGAGCTTGACCAAGTCCCAACAGAGTACACCATCACTCACTC 399
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleLeuValPheLeuThrLeu 140
Db 400 ATCGAGCAGAACTCCGGTTGGAAGTTGTATCCGTAACGGTAACATCGAGTGATCTTG 459
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
Db 460 CAGGACGTCACCGTAAGTACAGCTGCTGATCTGACACTCCGAGTCTCTGCCAC 519
QY 161 ProGlyTyr---AsnLysTrpPhePheValThrValThrAsnMetMetGlyAsnMet 179
Db 520 ACCGGTTACACCAACAAGTGTCTCTCGTACCATCACCAACATCATGGTTACATG 579
QY 180 LysIleThrIleAsnGlyLysLeuLeuAspThrIleLysValLysGluLeuThrGlyLe 199
Db 580 AAGTTGTATCATCAGCGGTGAGTTGAAGCAGTCCGAGAGATCGAGGACCTGGACGAGTC 639
QY 200 AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr 219
Db 640 AAGCTGGACAAGACCATCTCTCGGTATCGACGAGACATCGAC-----684
QY 220 SerAspSerAspAsnIleAsnMetTrpIleArgAspPheThrIlePheAlaLysGluLeu 239
Db 685 -----GAGAACCAAGATGTGTGGATCGTGTGACTTCAACATCTCTCCAAAGGAGCTG 735
QY 240 AspGlyLysAspIleAsnIlePheAsnSerLeuGlnTyrThrAsnValLysAsp 259
Db 736 TCCACGAGGACATCAATGCTGTACAGGGTCAGATCTCGAGAACGTCATCAAGGAC 795
QY 260 TyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeu 279
Db 796 TACTGGGTAAACCACTGAAGTTCACACCGAGTACTACATCATCAACGACAACTACATC 855
QY 280 AsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299
Db 856 GACCGTTATCATCCGCCAGAGTCAACGTCCTGTGTCTGTCTGTCTGTCTGTCTGTCT 915
QY 300 AspPheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThr 319
Db 916 AAGCTGTACACCGGTAAACCTATCATCAAGTCCGTCCTCGACAAAGACCCCTTACTCC 975
QY 320 ArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
Db 976 CGTATCTCTGAACGGTGACAACTATCATCTCGACATGCTGTGTACAACTCCCGTAACTAG 1035
QY 340 LeuPheMetLysAsnGluThrMetTyrAla-----AspAsnHisSerThrGluAsp 356
Db 1036 ATCATCCGTGACACCGACCATCTACCCACCGAGGTGTGTGTGTGTGTGTGTGTGTGT 1095
QY 357 IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln 376
Db 1096 GTCTACGCCCTGAAGCTGACGTCAACCTGGGTAACTACGAGTATCGGT---ATCTTCTCC 1152
QY 377 IleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsn 396
Db 1153 ATCAAGACATCTGTCTCAAGNACAAAGTACTGTCTCCAGATCTTC---TCTCTCTCCGT 1209
QY 397 GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGly 415
Db 1210 ---GAGAACACCATGCTGCTGGCGACATCTCAAGCCTTGGCGTTTCTCTCC-----1257
QY 416 AspTyrTrpArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeu 435
Db 1258 -----TTCAAGAACCGCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1305
```

```
QY 436 LeuGluSerThrSerThrHisTrpGlyPheVal-ProValSerGlu 450
Db 1306 CTGCTGTCCACTCCTCTCTTCTGGAAGTTATCTCTCCGTCACCCAG 1351
```

## RESULT 8

```
AAV30591
ID AAV30591 standard; DNA; 1469 BP.
AC AAV30591;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type D toxin C fragment gene in pETHisb.
XX
KW Antitoxin; vaccine; neurotoxin; toxin D; intoxication; immunogen;
KW botulism; BotD; ds.
XX
OS Clostridium botulinum serotype D.
OS Synthetic.
FH Key Location/Qualifiers
CDS 108..1463
/*tag= a
PN WO9808540-A1.
XX
XX 05-MAR-1998.
XX
XX 28-AUG-1997; 97WO-US15394.
XX
XX 28-AUG-1996; 96US-0704159.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
PA
PI Thalley BS, Williams JA;
XX
XX WPI; 1998-230234/20.
DR P-PSDB; AAW68398.
XX
XX Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX
XX Example 47; Page 352-53; 428pp; English.
XX
XX This is the DNA sequence of the Clostridium botulinum serotype D
XX neurotoxin fragment C gene contained in plasmid pETHisb. The
XX encoded BotD fragment C polypeptide (see AAW68398) has a His-tagged
XX N-terminal extension. The vector can be used to express native
XX soluble BotD C fragment in Escherichia coli host cells. The
XX invention relates to recombinant proteins derived from C. botulinum
XX toxins. Methods are provided which allow for the isolation of
XX soluble recombinant clostridial toxins free of significant
XX endotoxin contamination. Preferred hosts are E. coli, insect cells
XX and yeast cells. The recombinant toxins are used as immunogens for
XX the production of vaccines and antitoxins that are useful in the
XX treatment of humans and animals at risk of intoxication with
XX clostridial toxin.
XX
XX Sequence 1469 BP; 581 A; 146 C; 245 G; 497 T; 0 other;
```

```
Alignment Scores:
Pred. No.: 7,06e-63 Length: 1469
Score: 803.50 Matches: 186
Percent Similarity: 60.22% Conservative: 85
Best Local Similarity: 41.33% Mismatches: 151
Query Match: 33.61% Indels: 28
DB: 19 Gaps: 11
```

US-09-910-186A-10 (1-450) x AAV30591 (1-1469)

QY 4 PropheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleAsnGluTyr 23

```

1199 TGCATTAAATACAGAGTAATTTAGGTAAATTATGTTAGTATAGGT---ATATTAGTATAAA 1255
QY 378 nProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGly 398
Db 1256 AAATATTGTTATCTAAATAAATAATATGTTAGTCAAAATTTTC---TCTAGTTTATAGG---GA 1309
QY 398 uAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGlyAspTr 417
Db 1310 AAATACAAATGCTTACAGAGATATATATAAACTTGGAGATTTC----- 1355
QY 417 pTyrArgHisAsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGly 437
Db 1356 -TTTAAATAATCATACAGCGCAGTTCAGTAACT-----AATTATGAAACAAACTATT 1408
QY 437 uSerThrSerThrHisTyrGlyPheVal 446
Db 1409 ATCACTTCATCTTTTGGAAATTTATT 1436

RESULT 9
AAZ87215
ID AAZ87215 standard; DNA; 1400 BP.
XX
AC AAZ87215;
XX
DT 08-MAY-2000 (first entry)
XX
DE DNA encoding synthetic BoNT serotype E (BoNTE) Hc fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype E;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 10..1359
FT /tag= a
FT /product= "Synthetic botulinum neurotoxin serotype E
FT (BoNTE) heavy chain C-terminal fragment (Hc)"
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15570.
XX
PR 10-JUL-1998; 98US-0092416.
PR 12-MAY-1999; 99US-0133870.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
XX
DR P-PSDB; AAY77137.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
XX
XX
PS Disclosure; Page 43-44; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC serotypes A-G. In preferred embodiments of the invention, the vector is
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as

```

diagnostic tools for the diagnosis of botulism. The transformed host cells can be used to analyse the effectiveness of drugs and agents which inhibit toxin effects. The vaccine currently used against botulism is dangerous and expensive to produce, and contains formalin, which is very painful for the recipient. Also, the vaccine is incomplete, in that only 5 of the 7 serotypes are represented in the formulation. The novel vaccine overcomes these problems, as it is easily purified, and available in large quantities. It is also expressed in the lymph nodes for a better immune response. Sequences AA82712-787217 represent synthetic DNA sequences encoding PONT HC fragments used in the present invention. These were optimised for codon usage for expression in yeast.

SQ Sequence 1400 BP; 483 A; 298 C; 232 G; 387 T; 0 other;

Alignment Scores:

Pred. No.:	1.13e-43	Length:	1400
Score:	589.50	Matches:	152
Percent Similarity:	51.93%	Conservative:	90
Best Local Similarity:	32.62%	Mismatches:	159
Query Match:	24.65%	Indels:	65
DB:	21	Gaps:	17

US-09-910-186A-10 (1-450) x AA287215 (1-1400)

QY	2	ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn	21
DB	64	AGTATCTCTTTTAAAGCTTCTCTTATACAGATGATAAAATTTTAAATTCCTACTCAAC	123
QY	22	GlutyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr	41
DB	124	AGTTCTTCAGAGAAATTAACTCTTCTCCGTTTAAACATGAGATACCAAGATGATAAA	183
QY	42	LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlyAspValGlnLeuAsn	61
DB	184	TACGTCGACACTCCGGTTACGACTCCAATATCAACATTAAACGCTGAGCTGACAAGTAC	243
QY	62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle	81
DB	244	CCAACTACAAAAACCAATTC-----GGTACTCTACACGACACAGCTTACTGACTGAAC	297
QY	82	ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe	101
DB	298	ATCTCTCAAAACGACTACATTATCTACGACAAACAGATCAAGAAGTCTCTATTCTTCTTC	357
QY	102	TpIleArgIle-----AsnLysTyrpValSerAsnLeuProGlyTyrThrIle	117
DB	358	TGGTTCAGGATTCTTACTACGACACAGATCGTCACGGTTAACACGAGTACACATTC	417
QY	118	IleAspSerValLys-----AsnAsnSerGlyTyrpSerIleGlyIleIleSerAsnPheLeu	136
DB	418	ATCAACTGTATGAGAGACACAACTCCGGTTGGAAGGTCTCTTTAAACACACAGGATC	477
QY	137	ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle	156
DB	478	ATTTCGACCTTGCAGACACACGCGAGGTATTAAACCAAGTTAGCATTCACACTCGGTAAC	537
QY	157	SerAsnAsnAlaProGlyTyr-----AsnLysTyrpPhePheValThrValThrAsnMet	175
DB	538	GCAAAACGGTATTTCGACTACATCAACAAGTGGATTTTCGTCCTCATTCACTCAACGACAGA	597
QY	176	MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu	195
DB	598	TTAGTGTACTCTAGCTTTACATTACGGTACTTAATTCGACCAAAAGTCCTCATTTTAAAC	657
QY	196	LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr	215
DB	658	TTAGGTAACATTTCAGCTTCTCGACAACTCTATTCAAGATCGTTAACTGCAGTACACC	717
QY	216	GlyLeuIleThrSerAspSerAspAsnIleAsnMetThrIleArgAspPheTyrIlePhe	235
DB	718	AGATACATT-----GGCATTAGATACCTTCAACATTTC	750
QY	236	AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAspSerLeuGlnTyrThrAsn	255

Db	751	GACAGGAGTTAGACGAGACCGGAGATTCAAACTTTATACAGCAACGAACTTAACACCAAT	810
Qy	256	ValValIysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn	275
Db	811	ATTTTGAAGGACATCTGGGTAACACTTGCITTCAGACAAGTAATACTATTATAAC	870
Qy	276	IleAspTyrLeuAsnArgTyrMet-----TyrAlaAsnSerArgGlnIleValPheAsn	293
Db	871	TGTGTTAAAGCAACAACTTCATTGTATGAGGAAAGGATTCCTACTTTAAGCATTAACAC	930
Qy	294	ThrArgArgAsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIleIleLys	310
Db	931	ATCAGAAGCACTATCTTTAGCTACACATATACTCTGGTATCAAGTTAAGATCCAA	990
Qy	311	ArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe	329
Db	991	AGAGTTAAACAACCTCTCTACTAAGGATAACCTTGTGTAGAAAGACGATCAGTCTATATT	1050
Qy	330	AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla	349
Db	1051	AACTTCGTCGCTACAGAGACT-----CACTTATTC-----CCATATATAGCT	1092
Qy	350	AspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIle	369
Db	1093	GATACCGCTACCAACAAC-----AAGGAGAAGACCATCAAGATC	1131
Qy	370	AsnAsp-----AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAla	387
Db	1132	TCCTCCCTCGCAACAGATTTAACCAAGTCGCTGTATGAACCTCGTCGGTAACAATGT	1191
Qy	388	SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly	407
Db	1192	ACCATGAACCTTTAAATAAT---AATGGAATAAT-----ATTGGG	1230
Qy	408	ThrTyrArgPheArg-----LeuGlyGlyAspTrpTyrArgHisAsnTyrLeu	423
Db	1231	TTGTAGTTTCAAGCAGATACACTAGTTGGTACTAGTTGGTAT-----	1275
Qy	424	ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHis---	442
Db	1276	-----TATCCCATGACAGATCACACCACAGCAATGGA	1311
Qy	443	-----TrpGlyPheVal 446	
Db	1312	TGTTTTTGGAACTTTATT 1329	
RESULT	10		
AAA54488			
ID	AAA54488	standard; DNA; 1400 BP.	
XX	AAA54488;		
XX	11-APR-2001	(first entry)	
XX	Botulism toxin heavy chain C-terminal coding sequence (serotype E).		
XX	Botulism; toxin; neurotoxin; heavy chain; recombinant expression;		
KW	recombinant vector; antigen; immune response; vaccine; bacterium;		
KW	infection; ds.		
XX	Synthetic.		
OS	Clostridium botulinum.		
XX	Key	Location/Qualifiers	
FH	CDS	10..1358	
FT		/tag= a	
FT		/product= H_C peptide fragment	
XX	W0200067700-A2.		
XX	16-NOV-2000.		





```

Db      1312 TGTITTTGGAACTTTATT 1329

RESULT 11
AAZ867218
ID      AAZ867218 standard; DNA; 2532 BP.
XX
XX      AAZ867218;
XX
XX      08-MAY-2000 (first entry)
XX
XX      DNA encoding native botulinum neurotoxin serotype A (BoNTA).
DE
XX      Botulinum neurotoxin; heavy chain; BoNT; serotype A;
KW      Venezuelan equine encephalitis virus replicon;
VVE: botulism; vaccine; diagnosis; drug screening; ds.
XX
XX      Clostridium botulinum.
OS
XX
XX      Key      Location/Qualifiers
FH      CDS      1..2532
FT      /*tag= a
FT      /product= "Native botulinum neurotoxin serotype A
FT      (BoNTA) heavy chain"
FT      /transl_except= (pos:2509..2529, aa:Gly)
XX
XX      WO200002524-A2.
XX
XX      20-JAN-2000.
XX
XX      09-JUL-1999; 99WO-US15570.
XX
XX      10-JUL-1998; 98US-0092416.
XX      12-MAY-1999; 99US-0133870.
XX
XX      (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
XX      Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX      WPI; 2000-160827/14.
XX      P-PSDB; AAY77140.
XX
XX      Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
XX      toxin serotypes A-G, is used for inducing an immune response against
XX      botulinum -
XX
XX      Example 3; Page 49-50; 54pp; English.
XX
XX      The invention relates to novel vaccines that induce a protective immune
XX      response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
XX      and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
XX      DNA construct comprising a vector, and at least one nucleic acid
XX      fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
XX      serotypes A-G. In preferred embodiments of the invention, the vector is a
XX      Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this
XX      vector results in the production of large amounts of a protein encoded by
XX      a sequence cloned into the replicon. The constructs are used to produce
XX      vaccines against botulism. The proteins can also be used as diagnostic
XX      tools for the diagnosis of botulism. The transformed host cells can be
XX      used to analyse the effectiveness of drugs and agents which inhibit toxin
XX      effects. The vaccine currently used against botulism is dangerous
XX      and expensive to produce, and contains formalin, which is very painful
XX      for the recipient. Also, the vaccine is incomplete, in that only 5 of
XX      the 7 serotypes are represented in the formulation. The novel vaccine
XX      of overcomes these problems, as it is easily purified, and available in
XX      large quantities. It is also expressed in the lymph nodes for a better
XX      immune response. The present sequence represents DNA encoding native
XX      BoNTA heavy chain used in an exemplification of the present invention.
XX
XX      Sequence 2332 BP; 1038 A; 241 C; 404 G; 849 T; 0 other;

Alignment Scores:
Pred. No.: 4,32e-43 Length: 2532
Score: 586.50 Matches: 149

```

Percent Similarity:	48.94%	Conservative:	82
Best Local Similarity:	31.57%	Mismatches:	186
Query Match:	24.53%	Indels:	55
DB:	21	Gaps:	11
US-09-910-186A-10 (1-450) x AAZ87218 (1-2532)			
Qy	3	IleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleLeuAsnGlu	22
Db	1186	ATACCTTTTTCAGCTTTCCAAATACGTTAGATAATCAAGATTATATCTACATTTACGAA	1245
Qy	23	TyPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu	42
Db	1246	TATATTAAAGAAATATTATTAATCTCTATATGAATTAAGATATGAAGTAATCATTTA	1305
Qy	43	ValAspThrSerGlyTyThrAsnGlnGluValSerGluGluGlyAspValGlnLeuAsnPro	62
Db	1306	ATAGACTTATCTAGGTATGCATCAAAATATAATATGGTAGTAAGTAATTTTGATCCA	1365
Qy	63	IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys	79
Db	1366	ATAGATAAAAAATCAAATTCAAATATTATTAATTTAGAAAGTAGT-----AAA	1410
Qy	80	ValIleValThrGlnAsnGluAsnIleValTyRsnSerMetTyRgLuSerPheSerIle	99
Db	1411	ATTGAGGTAAATTTTAAAAAATGCTATGTATATATATGATGTATGAATAATTTTAGTACT	1470
Qy	100	SerPheTrpIleArgLysAsnLysTrpValSerAsnLeuPro-----GlyTyThr	116
Db	1471	AGCTTTTGGTAGAAGATTCCTAAGTATTTTAAACAGATATAAGCTAAATAATGAATATACA	1530
Qy	117	IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu	136
Db	1531	ATAATAATTGTATGGAAATATATTCAGAGATCGAAGATCACITTAATTTGGTGAATA	1590
Qy	137	ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyTrpAspIle	156
Db	1591	ATCTGGACTTTTACAGGATATCTCAGGAATATAAACAAAGAGTAGTTTTTAAATACAGCTCAA	1650
Qy	157	SerAsnAsnAlaProGlyTyR-----AsnLysTrpPhePheValThrValThrAsnAsnMet	175
Db	1651	ATGATTAATATATACAGATTATATAACAGATGGATTTTGTAACTATCATTAATAATAGA	1710
Qy	176	MetGlyAsnMetLysIleTyIleAsnGlyLysLeuIleAspThrIleLysValLysGlu	195
Db	1711	TTAAATAACCTCTAAATTTATATAATGGAAGATTATAGATCAAAACCAATTCGAAT	1770
Qy	196	LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr	215
Db	1771	TTAGGTAATATTCATCCTAGTAAATAATAAATGTTTAAATTAGATGGTTGTAGAGATACA	1830
Qy	216	GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyIlePhe	235
Db	1831	-----CATAGATATATTGGATAAAATATTATTAATCTTTT	1866
Qy	236	AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsn	255
Db	1867	GATAAGGAATTAATCAAAAGAAATCAAGATTATATGATAATCAATCAAAATTCAGGT	1926
Qy	256	ValValLysAspTyTrpGlyAsnAspLeuArgTyRsnLysGluTyTyMetValAsn	275
Db	1927	ATTTTAAAGACATTTTGGGTGATTTTACAAATATGATAACCACTATATGTGTAAT	1986
Qy	276	IleAspTyRLeuAsnArgTyRMetTyRAlaAsnSer-----	287
Db	1987	TTATATGATCCAAATAAATATGCGATGTGAAATAATATGAGGTATTAGAGGTATATGTAT	2046
Qy	288	-----ArgGlnIleValPheAsnThrArg--ArgAsnAsnAsnAspPheAsn	302
Db	2047	CTTAAAGGGCCTAGAGGTAGCGTAATGACTACAAACATTTATTAATTTCAAGTTGTAT	2106
Qy	303	GluGlyTyRLeuLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg	322

```

Alignment Scores:
Pred. No.:      4.32e-43
Score:          586.50
                Length: 2532
                Matches: 149

```

Db 2107 AGGGGCAAAATTTATATAAATAAATATGCTTCGGAAATAAAGATATATTTGTTAGA 2166  
 Qy 323 GlyGlyAspLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342  
 Db 2167 AATATGATCGTGATATATTAATAGTAGTAAATAAATAAAGATATAGGTTA----- 2220  
 Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
 Db 2221 -----GCTACTATGCAATCACAGCGCGGTAGAAAATACTAAGT 2262  
 Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleGlnProMetAsnAsn 382  
 Db 2263 GCATTAGAAATACCTGATAGTAGGAATCTAAGTCAAGTAGTAGTAATGAAGTCAAAAAT 2322  
 Qy 363 ThrTyrTyrAlaSerGlnIlePheLysSerAspPheAsnGlyGluAsnIleSerGly 402  
 Db 2323 GATCAAGAAATAACAATAAATAGC---AAATGAATTTACAAGATAATAATGGGAATGAT 2379  
 Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHisAsnTyr 422  
 Db 2380 ATAGGCTTTATAGGATTCATCAGTTI-----AATAT 2412  
 Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
 Db 2413 ATAGCTAACTAGTAGCAAGTAATTTGGTATAATAGACAAATAGAAAGATCTAGTAGGACT 2472  
 Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450  
 Db 2473 TTGGGTCTCATGGGAATTTATTCCTGATAGATGAT 2508  
 RESULT 12  
 AAT29244  
 ID AAT29244 standard; DNA; 3891 BP.  
 XX AC AAT29244;  
 XX DT 09-JUL-1996 (first entry)  
 XX DE C. botulinum type A neurotoxin gene.  
 XX KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen; ds.  
 XX OS Clostridium botulinum.  
 XX PN W09612802-A1.  
 XX PD 02-MAY-1996.  
 XX PF 23-OCT-1995; 95WO-US13737.  
 XX PR 07-JUN-1995; 95US-0480604.  
 XX PR 24-OCT-1994; 94US-0329154.  
 XX PR 16-MAR-1995; 95US-0405496.  
 XX PR 14-APR-1995; 95US-0422711.  
 XX PA (OPHI-) OPHIDIAN PHARM INC.  
 XX PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;  
 XX PI Williams JA;  
 XX DR WPI; 1996-230603/23.  
 XX DR P-PSDB; AAR95010.  
 XX PT Fusion proteins comprising non-toxin protein and part of toxin -  
 XX PT useful to form anti-toxins against Clostridium botulinum type A, and  
 XX PT C. difficile type toxins, and to treat C. difficile intoxication,  
 XX PT partic. diarrhoea  
 XX PS Example 22; Page 344-350; 434pp; English.  
 XX CC The Clostridium botulinum type A neurotoxin gene (AAT29244) has been  
 CC cloned and sequenced. It encodes a protein (AAR95010) that is  
 CC processed to form a dimer composed of a light and a heavy chain.

CC The 50 kDa C-terminal portion of the heavy chain, or C fragment  
 CC (AAR95008), was produced using a synthetic gene (AAT29245) having codon  
 CC usage altered to improve expression in *Escherichia coli*. Fusion  
 CC proteins of the type A toxin or C fragment, with e.g. maltose  
 CC binding protein or polyhistidine affinity tag, are used to generate  
 CC neutralising antitoxins and in vaccine compans.  
 XX SQ Sequence 3891 BP; 1580 A; 392 C; 629 G; 1290 T; 0 other;

## Alignment Scores:

Pred. No.:	7.27e-43	Length:	3891
Score:	586.50	Matches:	149
Percent Similarity:	48.94%	Conservative:	82
Best Local Similarity:	31.57%	Mismatches:	186
Query Match:	24.53%	Indels:	55
DB:	17	Gaps:	11

US-09-910-186a-10 (1-450) x AAT29244 (1-3891)

Qy	3	IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu	22
Db	2545	ATACCTTTTCAGCTTCCAAATACGTAGATTAATCAAGATTTATCTACATTTACTGAA	2604
Qy	23	TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu	42
Db	2605	TATATTAAGAATAATATTAATACTTCTATATTGATTTAAGATATCAAAAGTAATCATTTA	2664
Qy	43	ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro	62
Db	2665	ATAGACTTATCTAGTATGCATCAAAATAAATATTGGTAGTAAGTAATTTTGTATCCA	2724
Qy	63	IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys	79
Db	2725	ATAGATAAAATCAAAATTCATTTATTTAATTTAGAAAGTAGT-----AAA	2769
Qy	80	ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle	99
Db	2770	ATTGAGTAATTTTAAAAATGCTATTGTATATAATAGTAGTATGATGAAAATTTTAGTACT	2829
Qy	100	SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr	116
Db	2830	AGCTTTGGATAAGAATTCCTAAGTATTTTAAACAGTATAAGTCTAAATAATGAATATACA	2889
Qy	117	IleIleAspSerValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeu	136
Db	2890	ATAATAAATGTATGGAAATAATTCAGGATGGAAGTATCATCTTAATATGCGTGAATA	2949
Qy	137	ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle	156
Db	2950	ATCTGGACTTTACAGGATACTCAGGAAATAAAACAAAGAGTAGTTTTTAAATACAGTCAA	3009
Qy	157	SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet	175
Db	3010	ATGATTAATATATCAGATTATATAAACAAGATGGATTTTGTAACTATCATCTAATAATAGA	3069
Qy	176	MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu	195
Db	3070	TTAATAACTCTAAATTTATATAAATGGAAGATTATAAGATCAAAACCAATTTCAAT	3129
Qy	196	LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr	215
Db	3130	TTAGGTAATATTCATGCTAGTATAATAATGTTTAAATTTAGATGGTGTAGAGATACA	3189
Qy	216	GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe	235
Db	3190	-----CATAGATATATTGGATAAATATTATTTTCTTTT	3225
Qy	236	AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn	255
Db	3226	GATAGGAATAATAAGAAAAGAAATCAAGATTTTATATGATCAATCAATCAATCAGGT	3285
Qy	256	ValValLysAspTyrTrpGlyAsnAspLeuValGlyTyrAsnLysGluTyrTyrMetValAsn	275

Db 3286 ATTTTAAAGACTTTTGGGTGATATTATTAACAATATGATAAACCACATATATGTTAAAT 3345  
 Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer- 287  
 Db 3346 TTATATGATCAATAAATATGTCGATGTAATATGTTAGGTATTTAGAGGTATATGAT 3405  
 Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302  
 Db 3406 CTTAAAGGGCTAGAGGTAGCGTAATGACTACAAACATTTATTTAATTCAGATTGAT 3465  
 Qy 303 GluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322  
 Db 3466 AGGGGCAAAATTTATTAATAAAATATGCTCTGGAAATAAAGATATATTTGTTAGA 3525  
 Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342  
 Db 3526 AATAATGATCGTATATATTAATGATAGTTAAATAAATAAAGATATAGTTA----- 3579  
 Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
 Db 3580 -----GCTACTAATGTCATCACAGCGGCGTAGAAAAATACTAAGT 3621  
 Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
 Db 3622 GCATTAGAAATACCTGATGAGGAATCTAAGTCAAGTAGTATGATGACCAAAAT 3681  
 Qy 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
 Db 3682 GATCAAGGAATPACAAATAATATOC---AAATGAATTCAGAGATATAATGGAATGAT 3738  
 Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyAspTyrArgHisAsnTyr 422  
 Db 3739 ATAGGCTTTATAGGATTCATCAGTTI-----AATAAT 3771  
 Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440  
 Db 3772 ATAGCTAACTAGTAGCAAGTAATGTTATTAATAGACAAATAGAAAGATCTAGTAGACT 3831  
 Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450  
 Db 3832 TTGGGTTCTCATGGGAATTTATTCCTGTAGATGAT 3867  
 RESULT 13  
 AAC64582  
 ID AAC64582 standard; cdna; 4835 BP.  
 XX  
 AC AAC64582;  
 XX  
 DT 15-FEB-2001 (first entry)  
 XX  
 DE BoNT/A neurotoxin prototoxin nucleotide sequence SEQ ID NO:10.  
 XX  
 KW Human; procholecystokinin; CCK A receptor; CCK B receptor;  
 KW pancreatitis; antiinflammatory; ss.  
 XX  
 OS Clostridium botulinum.  
 XX  
 PN WO200061192-A2.  
 XX  
 PD 19-OCT-2000.  
 XX  
 PF 06-APR-2000; 2000WO-US09142.  
 XX  
 PR 08-APR-1999; 99US-0288326.  
 XX  
 PA (ALLR ) ALLERGAN SALES INC.  
 XX  
 PI Steward LE, Sachs G, Aoki KR;  
 XX  
 DR WPI; 2000-679416/66.  
 XX  
 PT New composition for treating acute pancreatitis, comprises a pancreatic  
 cell surface marker binding element, a translocation element that

PT transfers polypeptide across vesicular membrane, and a therapeutic  
 PT element  
 XX  
 PS Disclosure; Page 29-32; 50pp; English.  
 XX  
 CC The present invention describes a composition (I) for treating acute  
 CC pancreatitis. (I) comprises a first element containing a binding element  
 CC that binds to a pancreatic cell surface marker, a second element  
 CC containing a translocation element that facilitates polypeptide transfer  
 CC across the vesicular membrane, and a third element containing a  
 CC therapeutic element that inhibits enzyme secretion in pancreatic cell  
 CC cytoplasm. Also described is a method for making a therapeutic  
 CC polypeptide having a binding element selective for cholecystokinin (CCK)  
 CC receptor by expressing within a host cell a recombinant chimeric  
 CC polypeptide comprising an extein containing a therapeutic element and a  
 CC translocational element, and an intein located to the carboxy terminal  
 CC of extein having a cysteine, serine or threonine at its amino terminus,  
 CC and contacting the extein with a synthetic peptide comprising a CCK  
 CC amino acid sequence containing an amidated phenylalanine at a natural  
 CC C-terminus, and a cysteine, serine or threonine at its N-terminus, and  
 CC a nucleophilic reagent able to cause cleavage of the intein to form a  
 CC peptide bond between the extein C-terminus and synthetic peptide  
 CC N-terminus through the formation of an activated ester or thio ester  
 CC of pancreatic digestive enzymes, and prevents exocytic fusion of vesicles  
 CC containing secretory enzymes of pancreas. (I) is useful for treating  
 CC acute pancreatitis. the present sequence encodes the Clostridium  
 CC botulinum BoNT/A neurotoxin prototoxin which is given in the  
 CC exemplification of the present invention.  
 XX  
 SQ Sequence 4835 BP; 1934 A; 517 C; 756 G; 1628 T; 0 other;

Alignment Scores:  
 Pred. No.: 9,45e-43 Length: 4835  
 Score: 586.50 Matches: 149  
 Percent Similarity: 48.94% Conservative: 82  
 Best Local Similarity: 31.57% Mismatches: 186  
 Query Match: 24.53% Indels: 55  
 DB: 21 Gaps: 11

US-09-910-186A-10 (1-450) x AAC64582 (1-4835)

Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleLeuAsnGlu 22  
 Db 2902 ATACCTTTTCAGCTTTCCAAATACGTAGATAATCAAGATATTATCTACATTACTGAA 2961  
 Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42  
 Db 2962 TATATTAGAAATATTATTAATCTCTATTGAATTTAGATATGAAAGTAACTATTA 3021  
 Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62  
 Db 3022 ATAGACTTATCTAGGTATGCATCAAAATAAATATTGGTAGTAAAGTAAATTTGATCCA 3081  
 Qy 63 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79  
 Db 3082 ATAGATAAAATCAAAATCAATTTATTTATTTAGAAAGTAGT-----AAA 3126  
 Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99  
 Db 3127 ATTCAGGTAATTTTAAAAATGCTATTGTATATATAGTATGTATGATAATTTAGTACT 3186  
 Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116  
 Db 3187 AGCTTTTGGATAAGAATTCCTAAGTATTTTAAACAGATATAAGTCTAAATATGAATATACA 3246  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeu 136  
 Db 3247 ATAATAATGCTATGGAATAATATTCAGGATGGAAGTATCACTTAATATGCTGAATA 3306  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 3307 ATCTGGGCTTACAGGATCTACAGGAATAAATAAAGAGTAGTTTTTAAATACAGTCAA 3366



```

Db      3943  TGGAACTTTATT 3954
      ||| |||:::
RESULT 15
AAV30579
ID      AAV30579 standard; DNA; 3876 BP.
AC      AAV30579;
XX
XX      07-DEC-1998 (first entry)
XX
XX      Clostridium botulinum type B toxin gene from Danish strain.
XX
XX      Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW      botulism; BotB; ds.
XX
XX      Clostridium botulinum serotype B Danish strain.
OS
XX      WO9808540-A1.
XX
XX      05-MAR-1998.
XX
XX      28-AUG-1997; 97WO-US15394.
XX
XX      28-AUG-1996; 96US-0704159.
XX
XX      (OPHI-) OPHIDIAN PHARM INC.
PA
XX      Thalley BS, Williams JA;
PI
XX
XX      WPI; 1998-230234/20.
DR
XX      P-PSDB; AAW68392.
XX
XX      Host cell containing recombinant expression vector encoding
PT      Clostridium botulinum type B or E toxin - useful to treat humans
PT      and other animals at risk of intoxication with clostridial toxin
XX
XX      Example 35; Page 291-296; 428pp; English.
XX
XX      This is the coding region of the Clostridium botulinum serotype B
CC      (Danish strain) toxin gene that codes for a 1291-amino acid
CC      polypeptide (see AAW68392). The C fragment (see AAW68394) of the
CC      B toxin has been expressed as histidine-tagged protein in Escherichia
CC      coli host cells. The invention relates to C. botulinum recombinant
CC      toxin polypeptides. Methods are provided which allow for the
CC      isolation of soluble recombinant proteins free of significant
CC      endotoxin contamination. Preferred hosts for production of the
CC      recombinant proteins are E. coli, insect cells and yeast cells.
CC      The recombinant proteins are used as immunogens for the production
CC      of vaccines and antitoxins that are useful in the treatment of
CC      humans and animals at risk of intoxication with clostridial toxin.
XX
XX      Sequence 3876 BP; 1612 A; 370 C; 617 G; 1277 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 2,26e-42 Length: 3876
Score: 581.00 Matches: 161
Percent Similarity: 49.69% Conservative: 82
Best Local Similarity: 32.92% Mismatches: 168
Query Match: 24.30% Indels: 78
DB: 19 Gaps: 21
US-09-910-186A-10 (1-450) x AAV30579 (1-3876)
Qy      3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
      ::::::::::: |||||::: |||||::: |||||::: |||||:::
Db      2506  ATGCCGTTTCATCATATACCAATGATACAACTACTAATAAGAAATGTTTAAATAA 2505
      TTTT::: |||||::: |||||::: |||||::: |||||::: |||||:::
Qy      23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
      TTTT::: |||||::: |||||::: |||||::: |||||::: |||||:::
Db      2566  TATTAATAGCGAAATTTTAAATAATATATCTTTAAATTTTAAATATATAGGATATATATTTA 2629
      TTTT::: |||||::: |||||::: |||||::: |||||::: |||||:::
Qy      43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62

```

```

2626 ATAGATTATCAGGATATGGGCAAGGTAGAGGTATATGATCGAGTCGAGCTTAATGAF 2685
QY 53 IlePhePropheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysIleVal 82
Db 2686 -----AAAAATCAATTAATAATTAACAGTTCAGCA-----AATAGTAAGATTAGAGTG 2733
QY 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102
Db 2734 ACTCAAAATCAGATATCATATTAATTAAGTGTGTTCTTCATTTAGCGTTAGCTTTTGG 2793
QY 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116
Db 2794 ATAAGATACCTAAATATAAGATGATGATACAAATATATATTCATATGAATATACA 2853
QY 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
Db 2854 ATATTAATTGTATGAAAATATTAATTCGGGTGGAAATATCTATTAGGGGTATAGGATA 2913
QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
Db 2914 ATATGGACTTTAATTGATATAAATGGAATGGAACCAACCAATCGGTATTTTGAATATAACATA 2973
QY 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
Db 2974 AGAGAGATATATCAGAGTATATAAATAGATAGGTGTTTTCCTACTACTATAATAAT--- 3030
QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
Db 3031 TTGATAACGCTAAATTTATATTAATGTTAGCTAGCAATCAAAATACAGATATTAAGAT 3090
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 3091 ATAGAGAGATTATGCTAATGGTGAATATATTAATTAATGAT----- 3135
QY 216 GlyLeuIleThrSerAspSerAsnIleAsn---MetTrpIleArgAspPheTyrIle 234
Db 3136 -----GGTGATATAGATAGACACAACTTATTGATGAATATTTCACTATT 3183
QY 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThr 254
Db 3184 TTTAATACGGAATTAAGTCAATCAATTAATCAAGAAAGATATAAAATTCATATATAGC 3243
QY 255 AsnValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetVal 274
Db 3244 GAATATTTAAAGATTTTGGGGAATCCCTTAAATGACAAATAAGATATATATGTTT 3303
QY 275 AsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer-----ArgGlnIle 290
Db 3304 AATCGGGGAATAAAATTCATATATTAATACTAAAGAAAGATTCACCTGTAGTGGAATT 3363
QY 291 ValPheAsnThrArgArgAsnAsnAsn-----AspPheAsnGlu 303
Db 3364 TTACACGTACGAATATATCAAAATCTAAATATATAATAATATATAGAGATTATATATT 3423
QY 304 GlyTyrLysIleIleIleLysArgIleArgGlyAsnThr-----AsnAspThrArg 320
Db 3424 GGAGAAAAATTTATATAAAGA---AAGTCAAAATCTCAATCTATAAATGATGATATA 3480
QY 321 ValArgGlyGlyAspIleLeuTyrPheAsp---MetThrIleAsnAsnLys----- 336
Db 3481 GTAGAAAAGAGATTATATATCTAGATTGTTTAAATTAATCAAGAGTCGAGAGTA 3540
QY 337 ---AlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu 355
Db 3541 TATACCTATAAATATTTAAGAAAGAGAA-----GAA 3573
QY 356 AspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePhe 375
Db 3574 AAATGTTTTAGCTCCCTAATAGTCTGATGAGTTTACAAAT-----ACTATA 3624
QY 376 GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGln---IlePheLysSerAsn 394

```

```

Db 3625 CAATAAAGATATCATGAACCAACACATATAGTTGTCAGTGTGTTTTTAAAAA--- 3681
QY 395 PheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGly 414
Db 3682 ---GATGAAGAAAAGTACTGATGAGATAGGATTGATTGGTATTTCATCGTTTCTACGAATCT 3738
QY 415 Gly-----AspTrpTyrArgHisAsn 421
Db 3739 GGAATTGCTATTGAGAGATATAAGATTATTTTGTATAAGTAAATGGTAC----- 3789
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 3790 -----TTAAAGAGGTAAAAAGCAACCATATAATTTAAATTTG-----GGATGT 3834
QY 442 HisTrpGlyPheValProValSerGlu 450
Db 3835 AATGGCAGTTTATTCTCTAAAGATGAA 3861

```

Search completed: November 7, 2002, 17:51:59  
Job time : 279 secs

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	586.5	24.5	3891	1	US-08-480-604A-27	Sequence 27, Appl
2	586.5	24.5	3891	2	US-08-403-496A-27	Sequence 27, Appl
3	586.5	24.5	3891	4	US-08-913-136-27	Sequence 27, Appl
4	588.5	23.4	1402	1	US-08-480-604A-25	Sequence 25, Appl
5	558.5	23.4	1402	2	US-08-403-496A-25	Sequence 25, Appl
6	558.5	23.4	1402	4	US-08-913-136-25	Sequence 25, Appl
7	557.5	23.3	1330	1	US-08-480-604A-22	Sequence 22, Appl
8	557.5	23.3	1330	2	US-08-403-496A-22	Sequence 22, Appl
9	557.5	23.3	1330	4	US-08-913-136-22	Sequence 22, Appl
10	539.5	22.6	3509	4	US-09-255-829-19	Sequence 19, Appl
11	423.5	17.7	1858	1	US-08-668-381A-6	Sequence 6, Appl
12	420	17.6	1359	1	US-07-618-312A-1	Sequence 1, Appl

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:



```

; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
US-08-480-504A-27

```

```

Alignment Scores:
Pred. No.: 3,46e-53 Length: 3891
Score: 586.50 Matches: 149
Percent Similarity: 48.94% Conservative: 82
Best Local Similarity: 31.57% Mismatches: 186
Query Match: 24.53% Indels: 55
DB: 11 Gaps: 11
US-09-910-186A-10 (1-450) x US-08-480-604A-27 (1-3891)

```

```

Qy 3 IleProPheAsnIlePheSerThrAsnAsnSerLeuLeuLysAspIleleAsnGlu 22
Db 2545 ATACCTTTTCAGCTTCCAAATAGGTAGATAATCAAGATTATTACTACATTTACTGAA 2604
Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
Db 2605 TATATTAGAATATTATTAATCTCTATCTTCAATTTAAGATATCAAAAGTAAATCAATTTA 2664
Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62
Db 2665 ATAGACATTATCTAGGTATGCATCAAAATAATAATATTGGTAGTAAAGTAAATTTGTATCA 2724
Qy 53 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79
Db 2725 ATAGATAAAATCAAAATCAATTTATTTAATTTAAGAAAGTAGT-----AAA 2769
Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99
Db 2770 ATTGGGTAAATTTAAAAATGCTATTTCGTATATAATAGTATGTGAAAATTTTAGTACT 2829
Qy 100 SerPheThrPheLeuArgIleAsnLysTyrValSerAsnLeuPro-----GlyTyrThr 116
Db 2830 AGCTTTTGGATAAGAATTCCTAAGTATTTTAAACAGTATAAGTCTAAATGAATATACA 2889
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
Db 2890 ATATAAATTTGTATGGAAATAATATCAGATGGAAGTAGTACCTATTATGTGTGAATA 2949
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerThrAspIle 156
Db 2950 ATCTGGAGCTTTACAGGATCTACGAGAAATAAAACAAAGAGTAGTTTTTAAATACAGTCAA 3009
Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTyrPhePheValThrValThrAsnAsnMet 175

```

```

Db 3010 ATGATTAAATATATCAGATTATATAAATCAGATGATTTTGTAACTATCACTAATATAGA 3069
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
Db 3070 TTAATAAATCTCTAAATTTATATAAATGAAGATTATAGATCAAAACCAATTTCAAT 3129
Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 3130 TTAGGTAAATATTCATGCTAGTAATAATAATGTTTAAATTAGATGGTGTAGAGATACA 3189
Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235
Db 3190 -----CATAGATATATTGGATAAAATATTTTAATCTTTT 3225
Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn 255
Db 3226 GATAAGGAATTAATGAAAGAAATCAAGATTATATGATGATCAATCAATCAAGT 3285
Qy 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
Db 3286 ATTTTAAAGACATTTTGGGTGATTATTACAATATGATAAACCATACTATATGTTAAAT 3345
Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
Db 3346 TTATATGATCCAAATAATAATGTCGATGTAATAATAGTATAGGTATTAGAGTTATATGAT 3405
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302
Db 3406 CTTAAAGGGCTAGAGGTAGCGTAATGACTACAACATTTATTAAATCAAGTTTGTAT 3465
Qy 303 GluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322
Db 3466 AGGGGACAAAATTTATTATAAAAAATATGCTTCGAAATAAAGATAAATATTTGTTAGA 3525
Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
Db 3526 AATATGATCGTGTATATATTATGATAGTAGTTTAAAAATAAAGATATAGGTTA----- 3579
Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
Db 3580 -----GCTACTAATGCTATCATCACAGCAGGCGTAGAAAAATACTAAGT 3621
Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382
Db 3622 GCATGAATAACTGCTAGTATGAGGAATCTAAGTCAAGTAGTAGTATGAAGTCAAAAT 3681
Qy 383 ThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
Db 3682 GATCAAGGAATAACAAATAAATGCG---AAATGATTTTACAAGATAAATAATGGGAATGAT 3738
Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrPyrArgHisAsnTyr 422
Db 3739 ATAGGCTTTATAGATTTTCATCAGTTT-----AATAAT 3771
Qy 423 LeuValProThrValLysGlnGlyAsnTyrValSerLeuLeuGluSerThrSer----- 440
Db 3772 ATACATAACTAGTAGCAGCAAGTAAATGCTATATAATAGACAAATAGAAAGATCTAGTAGGACT 3831
Qy 441 -----ThrHisTrpGlyPheValProValSerGlu 450
Db 3832 TTGGTTTGCATCGGAATTTTCTCTGTAGATGAT 3867

```

```

RESULT 2
US-08-405-496A-27
; Sequence 27, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:

```

```

: ADDRESS: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94104
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US 08/405,496A
: FILING DATE: 16-MAR-1995
: CLASSIFICATION: 424
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/329,154
: FILING DATE: 25-OCT-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,907
: FILING DATE: 02-DEC-1993
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/985,321
: FILING DATE: 04-DEC-1992
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/429,791
: FILING DATE: 31-OCT-1989
:
: ATTORNEY/AGENT INFORMATION:
: NAME: INGOLIA, DIANE E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: OPD-01308
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
:
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3891 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3888
:
: US-08-405-496A-27
:
: Alignment Scores:
: Pred. No.: 3,46e-53 Length: 3891
: Score: 586.50 Matches: 149
: Percent Similarity: 48.94% Conservative: 82
: Best Local Similarity: 31.57% Mismatches: 186
: Query Match: 24.53% Indels: 55
: DB: 2 Gaps: 11
:
: US-09-910-186A-10 (1-450) x US-08-405-496A-27 (1-3891)
:
: Qy 3 lleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleleAsnGlu 22
: Db 2545 ATACCTTTTCAGCTTCCAAATACGTAGATAAACAAGATTATTACTACATTTACTGAA 2604
:
: Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleleuSerLeuGlnAsnArgLysAsnThrLeu 42
: Db 2605 TATATAAGATATATTAATCTTCTATCTATGAATTAAGATATGAAGTAAATCAATTTA 2664
:
: Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62
: Db 2665 ATAGACTTATCTAGTAGTCATCAAAAATAAATATGTTAGTAAAGTAAATTTTGATCCA 2724
:
: Qy 63 llePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79
: Db 2725 ATAGATAAAAAATCAAAATTAATTAATTTAGAAAGTAGT-----AAA 2769
:
: Qy 80 ValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIle 99

```

```

:
: Db 2770 ATTGAGCTAATTTTAAAAAATGCTATGTATATAAATAGTATGATGAAATTTTAGTACT 2829
:
: Qy 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116
: Db 2830 AGCTTTTGGATAAGAATTCCTAAGTATTTTAAACAGATAAGTCTAAATATGAATATACA 2889
:
: Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
: Db 2890 ATAATAATTGTAAGGAAATAATTCAGGATGGAAGATATCACTTAATATGGTGAATA 2949
:
: Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
: Db 2950 ATCTGGACTTTACAGGATACCTCAGGAATAAACAAGAGTAGTATTTTAAATACAGTCAA 3009
:
: Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
: Db 3010 ATGATTATATATCAGATTATATAACAGATGGATTTTGTAACTATCACTAATATATA 3069
:
: Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
: Db 3070 TTAATAACTCTAAATTTATATAAATGGAAGATTATAGATCAAAAACCAATTTCAAA 3129
:
: Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProaspThr 215
: Db 3130 TTAGGTAATATTCATGCTAGTAATAATAATATGTTAAATTAGATGTTGTAGAGATA 3189
:
: Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetIrpIleArgAspPheTyrIlePhe 235
: Db 3190 -----CATAGATATATTGGATAAAATATTTAATCTTTT 3225
:
: Qy 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnIlyThrAsn 255
: Db 3226 GATAAGGAATTAATGAAAGAAATCAAGATTATATGATTAATCAATCAATCAATCAGGT 3285
:
: Qy 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
: Db 3286 ATTTTAAAGACITTTGGGTGATTATTACATATGATAAACCATACTATATGTTAAT 3345
:
: Qy 276 IleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
: Db 3346 TTATATGATCCAAATAAATATGTCGATGTAATAATAGTAGTATTAGAGTTATATGTAT 3405
:
: Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302
: Db 3406 CTTAAAGGGCTAGAGGTAGCTAATGACTACAAACATTTATTTAATCAAGTTTGTAT 3465
:
: Qy 303 GluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322
: Db 3466 AGGGGACAAAAATTTATATAAAAAATATGCTCTCGAAATAAAGATAATATTTGTAGA 3525
:
: Qy 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
: Db 3526 AATAATGATCGTGATATATTAAGTAGTAGTTAAATAAAGAAATATAGGTTA----- 3579
:
: Qy 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
: Db 3580 -----GCTACTAATGCATCACAGCGAGCGTAGAAAAATACTACTAAGT 3621
:
: Qy 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPrometAsnAsn 382
: Db 3622 GCATTAGAAATACCTGAAGTAGGAAATCTAAGTCAAGTAGTAGTAAATGAAGTCAAAAT 3681
:
: Qy 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
: Db 3682 GATCAAGGAATAAACAATAAATGC---AAATGAATTTACAAGATAAATATGGGAATGAT 3738
:
: Qy 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyClyAspTyrPyrArgHisAsnTyr 422
: Db 3739 ATAGCTTTATAGAGATTTTCATCAGTTT----- 3771
:
: Qy 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
:
:

```



```
QY 303 GluGlyTyrLysIleIleLeuValArg 322
Db 3466 AGGGGACAAATTTATATAAAATATGCTCTCGAATAAAGTATATGTTAGA 3525
QY 323 GlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMet 342
Db 3526 AATAATGATCGTGTATATTAATAGTAGTGTAAAAATAAAGATATAGTTA----- 3579
QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
Db 3580 -----GCTACTAATCCATCAGCAGCGGTAGAGAAAATACTAAGT 3621
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382
Db 3622 GCATTAGAAATACCTGTAGTAGAATCTAAGTCAAGTAGTAGTAATGAAGTCAAAAAT 3681
QY 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
Db 3682 GATCAGGAAATCAACAAATAATGTC---AAATGAATTTACAAGATAATAATGGGAATGAT 3738
QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyr 422
Db 3739 ATAGGCTTTATAGGATTCATCAGTTT-----AATAAT 3771
QY 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 3772 ATAGCTAAACTAGTACGAAGTAATGGTATAATAGACAAATAGAAAGATCTAGTAGACT 3831
QY 441 -----ThrHisTrpGlyPheValProValSerGlu 450
Db 3832 TTGGGTGTCTCATGGGAATTTATTCCTGTAGATGAT 3867

RESULT 4
US-08-480-604A-25
; Sequence 25, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-480-604A-25

Alignment Scores:
Score: 8.44e-51 Length: 1402
Pred. No.: 558.50 Matches: 147
Percent Similarity: 48.73% Conservative: 83
Best Local Similarity: 31.14% Mismatches: 186
Query Match: 23.36% Indels: 56
Db: 1 Gaps: 11

US-09-910-186A-10 (1-450) x US-08-480-604A-25 (1-1402)
QY 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleAsnGluTyr 23
Db 45 CCATATCGAAGTCTCATATGGGTAGCATGCTGCTGCTACCTTCACTGAATA 104
QY 23 rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43
Db 105 CATCAAGAACAATCATCAATCACTCATCTGACCTGCTAGCAATCACTCATGAT 164
QY 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProil 63
Db 165 CGACCTGTCTCGTACGCTTCCAAAATCAACATCGGTCTTAAGTTAACTTCGATCGAT 224
QY 63 ePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLysVa 80
Db 225 CGACAAGAATCAGATCCAGCTGTTCAATCTGGAATCTTCC-----AAAAAT 269
QY 80 lIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSe 100
Db 270 CGAAGTTATCTGGAAGATGCTATCGTATACACTATGATGACGAAACATCTCCACCTC 329
QY 100 rPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThrIle 117
Db 330 CTTCTGGATCGGTATCCCGAAATACTTCACTCCATCTCTGAAACAAATGAATACACCAT 389
QY 117 eIleAspSerValLysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuVa 137
Db 390 CATCAACTGCATGGAAAACAATCTGTGGTGAAGATCTCTGAACTACGGTGAATCAT 449
QY 137 lPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSe 157
Db 450 CTGGACTCTGCAGGACACTCAGGAAATCAAAACAGCGGTGTTGTTATTAATCAATCTCAGAT 509
QY 157 rAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMe 176
Db 510 GATCAACATCTCTGACTACATCAATCGCTGGATCTCTGTTACCATCAACCAATCGTCT 569
QY 176 tGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLe 196
```

```

Db 570 GAATAACTCCAAATCTACATCAAGCGCGTCTGATCGACCAAGAACCGATCTCCAATCT 629
Qy 196 uThrGlyLeuAsnPheserLysThrPheGluLeuAsnLysIleProAspThrG1 216
Db 630 GGGTAACATCCACCTCTCTAATACATCATGTTCAACATGGAGCGTGTGCGACACAT-- 687
Qy 216 yLeuIleThrSerAspSerAspAsnIleAsnMetTTPileArgAspPheTyrIlePheAl 236
Db 688 -----CACGCGTCACTCGATCGATCAATCTTCAATCTGTGGA 725
Qy 236 aLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGluTyrThrAsnVa 256
Db 726 CAAGAAGTGAACGAAAGAAATCAAGACCTGTACGACACACCATCTCAATCTGGTAT 785
Qy 256 lValIysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnI1 276
Db 786 CCTGAAGACTTCGCGGTGACTACCTGCGAGTACGACAAACCGTACTACTGCTGAATCT 845
Qy 276 eAspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287
Db 846 GTACGATCCGAAACAAATACCTGACGTCAACAATAGTATCCGCGGTATACATGATCT 905
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAspPheAsnG1 303
Db 906 GAAAGGTCGCGTGTCTGTATGACTACACATCTACCTGAATCTCTCCGTGACCG 965
Qy 303 uGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgG1 323
Db 966 TGGTACCAAAATTCATCATCAAGAAATACGCGTCTGTTAAACAGGACAAATATGTTGCGAA 1025
Qy 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetIy 343
Db 1026 CAATGATCGTGTATACATCAATGTGTAGTATAGAACAAAGATACCGTCTG----- 1077
Qy 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363
Db 1078 -----GCTACCAATGCTTCTCAGCGTGTGTAGAAAGACATCTTGTCTGC 1121
Qy 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383
Db 1122 TCTGGAATCCGACGCTGTGTATCTGTCTCAGTGTGTGTATGAATGAATCCAAAGCA 1181
Qy 383 rTyrTyrThrAlaSerGlnIlePheLysSerAspPheAsnGlyGluAsnIleSerGlyI1 403
Db 1182 CCAGGGTATCACTAACAAATGC---AAATGAATCTGCAGGACAAACAATGGTAACGATAT 1238
Qy 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHisAsnTyrLe 423
Db 1239 CGGTTTCATCGGTTCCACCATGTC-----AACAAATAT 1271
Qy 423 uValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 1272 CGCTAAACTGGTCTCTCCAACTGGTACAAATCGTCAGATCGAAGCGTCTCTCGCACTCT 1331
Qy 441 ---ThrHisTrpGlyPheValProValSerGlu 450
Db 1332 GGGTGTCTTGGGAGTTCATCCGCGGTGATGAC 1365

```

## RESULT 5

US-08-405-496A-25

Sequence 25, Application US/08405496A

Patent No. 5919665

GENERAL INFORMATION:

APPLICANT: WILLIAMS, JAMES A.

TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM

TITLE OF INVENTION: NEUTROTOXIN

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: MEDLEN &amp; CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/405,496A

FILING DATE: 16-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01308

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 1402 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1386

US-08-405-496A-25

## Alignment Scores:

Pred. No.:	8,44e-51	Length:	1402
Score:	558.50	Matches:	147
Percent Similarity:	48.73%	Conservative:	83
Best Local Similarity:	31.14%	Mismatches:	186
Query Match:	23.36%	Indels:	56
DB:	2	Gaps:	11

US-09-910-186A-10 (1-450) x US-08-405-496A-25 (1-1402)

Qy	4	ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLysAspIleIleAsnGluTyr	23
Db	45	CCATATCGAAGTTCGTCATATGCTAGCATGGCTGCTGCTGCTACCTTCACCTGAATA	104
Qy	23	rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa	43
Db	105	CATCAAGACATCATCAATACCTCCATCTCTGAACTGCGCTACGAATCCAATCCTGAT	164
Qy	43	lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProil	63
Db	165	CGACCTGCTCTCGCTACGCTTCCAAATCAACATCGGTTCTTAAGTTAACTTCGATCCGAT	224
Qy	63	ePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLysVa	80
Db	225	CGACAAGATCAGATCCAGCTGTTCATCTGGAATCTCC-----AAAT	269
Qy	80	lIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSe	100
Db	270	CGAAGTTATCTCTGAAGAAATGCTATCTATACACTCTATGTACGAAACTCTCCACCTC	329
Qy	100	rPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThrI1	117

Db 330 CTTCTGTCCTCCGATATCCCGAATACTTCAACTCCATCTCTCTGACCAATGAATACACCAT 389  
 QY 117 etleaspSerVallyAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPhelLeuVa 137  
 Db 390 CATCACTGCTGATGGAACAATCTGTGTGTAAGAAATCTCTGAACACACGTTGAAATCAT 449  
 QY 137 lPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheserTyrrAspIleSe 157  
 Db 450 CTGGACTCTGCGAGGACATCAGGAATCAACACAGCGTGTGTATCAAAATCACTCTCAGAT 509  
 QY 157 rAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValrAsnAsnMetMe 176  
 Db 510 GATCAACATCTCTGACTACATCTGCTGATCTGTTACCAATCAACCAACATCTGCT 569  
 QY 176 tGlyAsnMetLysIleTyrrIleAsnGlyLysLeuIleAspThrIleLysVallyGluLe 196  
 Db 570 GAATACATCCAAATCTACATCAACGCGCTGTATGACCAAGAACGATCCCAATCT 629  
 QY 196 uThrGlyIleAsnPheserLysThrIlePheGluIleAsnLysIleProAspThrGl 216  
 Db 630 GGTAAACATCCACGCTTCTAATACATCATCTTCAAACTGGAGCGTGTCTGACACT-- 687  
 QY 216 yLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrrIlePheAl 236  
 Db 688 -----CACCGCTACATCTGTGATCAAAATCACTCAATCTGTGTGTA 725  
 QY 236 aLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrrThrAsnVa 256  
 Db 726 CAAGAAGACTGACGAAAGAAATCAAGACCTGTACGACACACCGTCCCAATCTGTGTAT 785  
 QY 256 lVallyAspTyrrTrpGlyAsnAspLeuArTyrrAsnLysGluTyrrMetValAsnIl 276  
 Db 786 CCGTAAAGACTTCTGGGTGACTACTCTGACAGCAACCGTACTACATCTGTGATCT 845  
 QY 276 eAspTyrrLeuAsnAlaGlyMetTyrrAlaAspSer----- 287  
 Db 846 GTACGATCCGACAAATACGTTGACGTCAACATGTAGTATCCGCGGTACATGTACCT 905  
 QY 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsnGl 303  
 Db 906 GAAAGGTCGCGTGTCTGTATGACTACCAACATCTACCTGAACCTCTCCCTGTACCG 965  
 QY 303 uGlyTyrrIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323  
 Db 966 TGTGTACCAATTCATCATCATCAAGATATAGCGCTGTGTACAGGACCAATCTGTTCCAA 1025  
 QY 323 yGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsnLeuPheMetLy 343  
 Db 1026 CAATGATCGTGATACATCAATGTGTAGTTAAGAAACAAAGAAATACCGTCTG 1077  
 QY 343 sAsnGluThrMetTyrrAlaAspAsnHisSerThrGluAspIleTyrrAlaIleGlyLeuAr 363  
 Db 1078 -----GCTACCAATGCTTCAGGCTGGGTAGAAAGATCTGTCTGTC 1121  
 QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383  
 Db 1122 TCTGGAATCCCGGAGTGTGTATCTCTCAGTGTGTATGAATCAACCAAGAACGA 1181  
 QY 383 rTyrrTyrrAlaSerGlnIlePheLysSerAspPheAsnGlyGluAsnIleSerGlyIl 403  
 Db 1182 CCAGGATATCATCAACAATGCG---AAATGAATCGCAGGACCAACATGTTACAGTAT 1238  
 QY 403 eCysSerIleGlyTyrrArgPheArgLeuGlyGlyAspTrpTyrrArgHisAsnTyrrLe 423  
 Db 1239 CGGTTTCATCGGTTCCACGAGTTC-----AACATAT 1271  
 QY 423 uValProThrVallyGlnGlyAsnTyrrAlaSerLeuLeuLeuSerThrSer----- 440  
 Db 1272 CGCTAAACTGGTGTCTCCAACTGATCAAAATCGTCAATCGATCGACGATCTCTCGCACTCT 1331  
 QY 441 ---ThrHisTrpGlyPheValProValSerGlu 450  
 Db 1332 GGGTGTCTCTGGGAGTTCATCCCGTGTGATGAC 1365

RESULT 6

US-08-915-136-25  
 ; Sequence 25, Application US/08915136  
 ; Patent No. 6290960  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KINK, JOHN A.  
 ; APPLICANT: THALLEY, BRUCE S.  
 ; APPLICANT: PADHYE, NISHA V.  
 ; APPLICANT: FIRCA, JOSEPH R.  
 ; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
 ; PREVENTION OF C. DIFFICILE DISEASE  
 ; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
 ; PREVENTION OF C. DIFFICILE DISEASE  
 ; NUMBER OF SEQUENCES: 32  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MEDLEN & CARROLL, LLP  
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200  
 ; CITY: SAN FRANCISCO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: UNITED STATES OF AMERICA  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/915,136  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/480,604  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/405,496  
 ; FILING DATE: 16-MAR-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/329,154  
 ; FILING DATE: 25-OCT-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/161,907  
 ; FILING DATE: 02-DEC-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/985,321  
 ; FILING DATE: 04-DEC-1992  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/429,791  
 ; FILING DATE: 31-OCT-1989  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: INGOLIA, DIANE E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPHD-01763  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 25:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1402 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1386  
 ; US-08-915-136-25  
 Alignment Scores:  
 Pred. No.: 8.44e-51 Length: 1402  
 Score: 558.50 Matches: 147  
 Percent Similarity: 48.73% Conservative: 83  
 Best Local Similarity: 31.14% Mismatches: 186  
 Query Match: 23.36% Indels: 56



```
? ZIPPED FILE NAME:
? CURRENT REMOVABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/405.496A
? FILING DATE: 16-MAR-1995
```



[illegible][illegible][illegible]

APPLICANT:	STAFFORD, DOUGLAS C.
TITLE OF INVENTION:	VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE





; GENERAL INFORMATION:  
 ; APPLICANT: Brown, Robert H.  
 ; APPLICANT: Fishman, Paul S.  
 ; APPLICANT: Francis, Jonathan W.  
 ; APPLICANT: Hosler, Betsy A.  
 ; TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN  
 ; TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fish & Richardson P.C.  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: USA  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/668,381A  
 ; FILING DATE: 21-JUN-1996  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/000,473  
 ; FILING DATE: 23-JUN-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Clark, Paul T.  
 ; REGISTRATION NUMBER: 30,164  
 ; REFERENCE/DOCKET NUMBER: 00786/269001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617/542-5070  
 ; TELEFAX: 617/542-8906  
 ; TELEX: 200154  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1858 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA  
 ; US-08-668-381A-6

Alignment Scores:  
 Pred. No.: 4,08e-36 Length: 1858  
 Score: 423.50 Matches: 133  
 Percent Similarity: 44.16% Conservative: 90  
 Best Local Similarity: 26.34% Mismatches: 165  
 Query Match: 17.71% Indels: 117  
 Gaps: 21

US-09-910-186a-10 (1-450) x US-08-668-381A-6 (1-1858)

QY 8 PheSerThrAsnAsnSerLeuLeuLysAspIleleAsnGluTyrPheAsnAsnIle 27  
 Db 491 TTTTCTTATCTCTAAAT-----CTGGATTGTTGGGTGATAATGAA 532  
 QY 28 AsnAsp-----SerLysIleLeuSerLeuGlnAsnArgLysAsn 40  
 Db 533 GAAGATATAGATGTATATTTAAAGAGTACAAATTTAAATTTAGATATATATATGAT 592  
 QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60  
 Db 593 ATTATATCATGATATATCGGGTTTAAATTCATCTGTAATAACATATCCAGATGCTCAATTG 652  
 QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGlu----- 75  
 Db 653 GTGCC-----GGAATAATGCGAAAGCAATACATTTAGTA 688  
 QY 76 -----AspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMet 93  
 Db 689 ACATGATCTCTCTGAAGTTATAGTGCATAAAGCTATGATATATGATATATGATATG 748

QY 94 TyrGluSerPheSerIleSerPheThrIleArgIleAsnLys---TyrValSerAsnLeu 112  
 Db 749 TTTAATAATTTTACCGTTAGCTTTTGGTTGAGGTTCCCTAAAGTATCTGCTAGTCATTTA 808  
 QY 113 Pro-----GlyTyrThrIleIleAspSerValLysAsnAsn----- 124  
 Db 809 GAACAATATGCGACAAATGAGTATTCATTAATTAGCTCTATGAAGAAACATAGCTCTATCA 868  
 QY 125 -----SerGlyTyrPheSerIleGlyIleIleSerAspPheLeuValPheThrLeuLysGln 142  
 Db 869 ATAGGATCTGTTGGAGTGTACCTTTAAAGTAAATAACTAATATGGACTTTTAAAGAT 928  
 QY 143 AsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGly 162  
 Db 929 TCCCGGGGAGAGTTAGACAAATAACTTTTAGG--GATTACCTGATAAATTTAATGCT 985  
 QY 163 Tyr-----AsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180  
 Db 986 TATTAGCAATAAATGGGTTTATTAACATTACTAATATGATAGATATTCTCTGCTAAT 1045  
 QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200  
 Db 1046 TTGTATATAATGGAGTACTTATGGGAAGTCAGAAATTACTGGTTTAGGAGCTATTAGA 1105  
 QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220  
 Db 1106 GAGGATAATAATAACATTAACAACTAGATAGA-----TGT 1141  
 QY 221 AspSerAspAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240  
 Db 1142 AATATATATATCAATACGTTCTATTGATAAATTTAGGATATTTGCCAAGCATTAAT 1201  
 QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyr 260  
 Db 1202 CCAAAAGAGATTGAAAAATTATACACAAGTTATTTATCTATATAACCTTTTAAAGAGACTTC 1261  
 QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276  
 Db 1262 TGGGAAACCCCTTTACGATATGATACAGAATATTTTAAATACCAAGTCTTCTAGTTCT 1321  
 QY 277 -----AspTyrLeu-----AsnArgTyrMetTyr 284  
 Db 1322 AAAGATGTTCAATTGAAAAATATAACAGATTATATGTTTGCATAATGCGCATCGTAT 1381  
 QY 285 AlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGly 304  
 Db 1382 ACTAACGGAAAAATGAAATATATATTAGAGGTTATATAAT-----GGA 1426  
 QY 305 TyrLysIleIleLysArgIleArgGlyAsnThrAsn---AspThrArgValArgGly 323  
 Db 1427 CTAAATTTATATAAAGATATACACCTAATAAGTAATAGATCTTTTGTAAATCA 1486  
 QY 324 GlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyr----- 338  
 Db 1487 GGTGATTTTATAATATATGTTATCATATAACAATAATGACACATCTGAGTTATCCG 1546  
 QY 339 -----AsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu 355  
 Db 1547 AAAGATGGAATGCTTTAATACTTGTAGATAGTCTAAGAGTAGTATATATGCCCA 1606  
 QY 356 AspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePhe 375  
 Db 1607 GGTATCCCTCTTTATAAAAAATGGAAGCAGTAAAA----- 1642  
 QY 376 GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPhe 395  
 Db 1643 -----TTGCGTGATTTAAAAACCTATCTGTACAACCTT-----AAATATAT 1684  
 QY 396 AsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGly 415  
 Db 1695 GATGATAAAATGTCATCT---TTAGACTAGTAGTACCCATAATGTCATAATGATGCAAC 1741

```

Qy 416 Asp-----TptTyrArgHisAsnTyrLeuValPro 425
Db 1742 GATCCAATAGGATATATTAATGCAACACTGGTAC---TTTAATCAATTTAAAGAT 1798

Qy 426 ThrVallylsGlnGlyAsnTyrAlaSerLeuLeuGluSerThrHisTrpGlyPhe 445
Db 1799 AAAATT-----TTAGGTGTGATGGTACTTT 1825

Qy 446 ValProValSerGlu 450
Db 1826 GTACCTACAGATGAA 1840

RESULT 12
US-07-618-312A-1
; Sequence 1, Application US/07618312A
; Patent No. 5389540
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr. Andrew J
; APPLICANT: Romanos Dr. Michael A
; APPLICANT: Clare Dr. Jeffrey J
; APPLICANT: Fairweather Dr. Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 14th Floor
; STREET: 2200 Clarendon Boulevard,
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618,312A
; FILING DATE: 19910516
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crawford Mr. Arthur R
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 510-51
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 0101 703 8750400
; TELEFAX: 0101 703 5253468
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
US-07-618-312A-1

Alignment Scores:
Pred. No.: 6.26e-36 Length: 1359
Score: 420.00 Matches: 131
Percent Similarity: 44.20% Conservative: 90
Best Local Similarity: 26.20% Mismatches: 169

```

```

Query Match: 17.57% Indels: 110
DB: 1 Gaps: 20

US-09-910-186A-10 (1-450) x US-07-618-312A-1 (1-1359)

Qy 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsn 25
Db 7 AATCTGGATTGTGGTTCATATGAAGAAGATATAGATTATATA----- 51

Qy 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45
Db 52 ---TTAAAAAGAGTACAAATTTTAAATTAGATATATAATATGATATTAATATCGATATA 108

Qy 46 SerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65
Db 109 TCTGGGTTTAATTCATCTGTAATAACATATCCAGATGCTCAATTGGTGCCC----- 159

Qy 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78
Db 160 -----GGAATAAATGGCAAGCAANTACATTTAGTAAACAATGAATCTTCT 204

Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98
Db 205 GAAGTTATATGTCATAAGCTATGATATTGAATATAATGATATGTTTAATAATTTTACC 264

Qy 99 IleSerPheTrpIleArgIleAsnLys---TtpValSerAsnLeuPro----- 113
Db 265 GTTAGCTTTGGTTGAGGTTCTCTAAAGTATCTGCTAGTCATTAGAACATAATGGCACA 324

Qy 114 ---GlyTyrThrIleIleAspSerValLysAsnAsn-----SerGlyTrp 127
Db 325 AATGAGTATTCAATAATTAGCTCTATGAAAAAACATAGCTCTATCAATAGGATCTGTTGG 384

Qy 128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147
Db 385 AGTGATCACTTAAAGTAACTAATTAATAGGACTTTAAAGATTCCTCGGAGGAGAAATT 444

Qy 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr-----AsnLys 165
Db 445 AGCAATAACTTTTAG---GATTACCTGATAATTTAATGCTTATTAGCAATAAA 501

Qy 166 TrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly 185
Db 502 TGGGTTTTTATACTATTACTAATGATAGTATCTCTGCTAAATTTGTTATATAATGA 561

Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205
Db 562 GTACTTATGGGAAGTCGCAAAATTAAGGAGCTATTAGAGAGGATATAATAATA 621

Qy 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIle 225
Db 622 ACATTAATAACTAGATAGA-----TGTAATAATAATAATAACAA 657

Qy 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245
Db 658 TACGTTTCTATTGATAAATTTAGGATATTTGCCAAGCATTAATCCCAAGAGATTGAA 717

Qy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu 265
Db 718 AATTATACACAAGTTTATTTATCTATAACCTTTTAAAGAGACTTCTGGGAAACCCCTTA 777

Qy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276
Db 778 CGATATGATACAGAATATTATTTAATACAGCAGTCTCTAGTTCCTAAAGATGTTCAATTG 837

Qy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289
Db 838 AAAAATAACAGATTATATGTTTGGCAAAATGCGCCCATCGTATACTAACGGAAATTTG 897

Qy 290 IleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrIleIleIle 309
Db 898 AATATATATTATAGAAGGTTATATAAT-----GGACTAAATTTTATTATA 942

```



```
QY 226 AsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspClyLysAspIleasn 245
Db 658 TACGTATCCATCGAAGTTCCTGATCTTCTGCAAGACACTGAACCCGAAGAGATCGAA 717
QY 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu 265
Db 718 AAACGTATACCACTACCTGCTATACACCTCTCTCGTGACTTCTGGGTAAACCGCTG 777
QY 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276
Db 778 CGTTACGACCGCAATATACCTGATCCCGGTAGCTTCTAGTCTTAAGACGTTTCAGGTG 837
QY 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289
Db 838 AAAACATCATCTACTACTACTCTACCTGACCAACGCGCTCTACACTAACGGTAACTG 897
QY 290 IleValPheAsnThrArgAsnAsnAsnAspPheAsnGluGlyTyrIleIleIle 309
Db 898 AACATCTACTACCGACGCTGTACAAC-----GGCCTGAATTCATCATC 942
QY 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyGlyAspIleLeuTyr 328
Db 943 AAACGCTACACTCCGAACAACGAATCTCTTCTGTTAAATCTGGTGACTCATCAA 1002
QY 329 PheAspMetThrIleAsnAsnLysAlaTyr-----AsnLeu 340
Db 1003 CTGTACGTTTCTTACACAACAACGACATCGTTGTTACCCGAAAGACGGTAACGCT 1062
QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1063 TTCAACAACCTGGACAGATCTCGCTGTTGGTTACACGCTCCGGGTATCCCGCTGTAC 1122
QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleGlnProMet 380
Db 1123 AAAAAAATGGAAGCTGTAA-----CTG 1146
QY 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1147 CGTGACCTGAAACCTACTCTGTAGCTG-----AACTGTAGACGACAAAACGCT 1200
QY 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAsp----- 416
Db 1201 TCT---CTGGGTCTGGTGTGTGATCCCAACGCTCAGATCGGTAAACCGCAACCGTGAC 1257
QY 417 -----TrpTyrArgHisAsnTyrLeuValProThrValLysGlnGly 430
Db 1258 ATCTGTATGCTTCTACTGCTGAC---TTCACCCACTCTGAAAGACAAATC----- 1305
QY 431 AsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu 450
Db 1306 -----CTGGGTGCGACTGCTACTCTGTTCCGACCGATCAA 1341
RESULT 14
US-08-280-228-1
; Sequence 1, Application US/08280228
; Patent No. 5571694
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; STREET: 1100 No. 5571694th Glee Road
; CITY: Arlington,
; STATE: Virginia,
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,228
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/618,312
; FILING DATE: 27-NOV-1990
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mary J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-08-280-228-1
Alignment Scores:
Pred. No.: 6,26e-36 Length: 1359
Score: 420.00 Matches: 131
Percent Similarity: 44.20% Conservative: 90
Best Local Similarity: 26.20% Mismatches: 169
Query Match: 17.57% Indels: 110
DB: 1 Gaps: 20
US-09-910-186a-10 (1-450) x US-08-280-228-1 (1-1359)
QY 6 AsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsn 25
Db 7 AATCTGGATTGTGGTTCATTAATGAAGAGATATAGATGTTATA----- 51
QY 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45
Db 52 ---TTAAAAAGAGTACAAATTTTAAATTTAGATATATAATTAATGATTAATATCGATATA 108
QY 46 SerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65
Db 109 TCTGGGTTTAATTCATCTCTATATACATATCCAGATGCTCAATTCGTGCC----- 159
QY 66 PheAspPheLysLeuGlySerSerGlyGlu-----AspArgGly 78
Db 160 -----GGAATAAATGCAAGCAATACATTTTAGTAACAATGAATCTTCT 204
QY 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98
Db 205 GAAGTTATAGTCATAAAGCTATGGATTTGAATATATATGATATGTTTAAATAATTACC 264
QY 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro----- 113
Db 265 GTTAGCTTTTGGTGGGTTCTCTAAAGTATCTGCTAGTCAATTAGAACAAATATGGCACA 324
```

Qy 114 ---GlyTyrThrIleAspSerValIysAsnAsn-----SerGlyTyr 127  
 |||:||||| |||:|||| |||:||||  
 Db 325 AATGAGTATTCAATAATAGTCTCTATGAAACACATAGTCTATCAATAGGATCTGGTTGG 384  
 |||:||||| |||:||||| |||:|||||  
 Qy 128 SerIleGlyIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147  
 |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 385 AGTGTATCAGTAAAGGTATACTTAATATGAGCTTTAAAGATCCCGGGGAGAGTT 444  
 |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr-----AsnLys 165  
 |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 445 AGCAAAATACCTTTTAGG--GATTACCTGATAAAATTAATGCTTATTAGCAATAAA 501  
 |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 166 TrpPhePheValThrValThrAsnMetMetClyAsnMetLysIleTyrIleAsnGly 185  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 502 TGGGTTTATTAACATACTACTAATGATAGATTACTCTCTAATTGATATAATAATGA 561  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 562 GTACTTATGGAGTGCAGAAATTAAGTCTTTAGGAGCTATTAGAGGATAATAATATA 621  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIle 225  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 622 ACATTAACCTAGTAGA-----TGTAATAATAATAATCAA 657  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 226 AsnMetIrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 658 TACGTTTCTATTGATAAATTAGGATATTTTGCAGCATTAATCCAAAGAGATTGAA 717  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu 265  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 718 AAATTATACACAGTATTATCTATCTATACCTTTTAAAGAGACTCTGGGGAACCCCTTTA 777  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle----- 276  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 778 CGATATGATACAGATATTATTTAATACAGTAGCTTCTAGTCTTAAGATGTTCAATTG 837  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgLys 289  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 838 AAAAATAACAGATTATATATTGACAAATCCGCCATFCGTATACTAACGGAATAATG 897  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 290 IleValPheAsnThrArgAsnAsnAspPheAsnGluGlyTyrLysIleIleIle 309  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 898 RATATATTATTAGAGGTTATATAAT-----GGACTAAATTTATTATA 942  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyLysAspIleLeuTyr 328  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 943 AAAAGATATACACCTAATAATGAATAGATCTTTTGTAAATCAGGTGATTTTATAAA 1002  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 329 PheAspMetThrIleAsnAsnLysAlaTyr-----AsnLeu 340  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 1003 TTATATGTATCATATACATAATAGCACATGTAGTTCGCCGAAAGATGGAATGCC 1062  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 1063 TTTAATAATCTGTAGAAATCTAAGAGTAGTATGATTAATGCCCGGTATCCCTCTTTAT 1122  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 1123 AAAAATAAGGAGCAGTAAA-----TTG 1146  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 381 AsnAsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyLysAsnIle 400  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 1147 CGTGATTTAAAAACCTATTCTGTACAACTT-----AAATTATATGATATAAAATGCA 1200  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 401 SerGlyIleCysSerIleGlyTyrTyrArgPheArgLeuGlyGlyAsp----- 416  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 1201 TCT---TTAGGACTAGTAGGTACCCATAATGGCAACGATCCCAATAGGAT 1257  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Qy 417 -----TrpTyrArgHisAsnTyrLeuValProThrValLysGlnGly 430  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 1258 ATATTATTGCAAGCACTGTAC---TTTAATCAITTAAGATAAAT----- 1305  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

Oy 431 AsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheValProValSerGlu 450  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||  
 Db 1306 -----TTAGGATGTGATTGGTACTTTGTACCTACAGATGAA 1341  
 |||:||||| |||:||||| |||:||||| |||:||||| |||:|||||

## RESULT 15

US-08-280-228-3  
 ; Sequence 3, Application US/08280228  
 ; Patent No. 5571694  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Makoff Dr, Andrew J  
 ; APPLICANT: Romanos Dr, Michael A  
 ; APPLICANT: Clare Dr, Jeffrey J  
 ; APPLICANT: Fairweather Dr, Neil F  
 ; TITLE OF INVENTION: VACCINES  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: NIXON & VANDERHYE P.C.  
 ; STREET: 1100 No. 5571694th Glebe Road  
 ; CITY: Arlington,  
 ; STATE: Virginia  
 ; COUNTRY: U.S.A.  
 ; ZIP: 22201-4714  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/280,228  
 ; FILING DATE: 25-JUL-1994  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/618,312  
 ; FILING DATE: 27-NOV-1990  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: GB 8926832.0  
 ; FILING DATE: 28-NOV-1989  
 ; APPLICATION NUMBER: GB 9006097.1  
 ; FILING DATE: 17-MAR-1990  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Wilson, Mary J  
 ; REGISTRATION NUMBER: 32,955  
 ; REFERENCE/DOCKET NUMBER: 117-163  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (703) 816-4000  
 ; TELEFAX: (703) 816-4100  
 ; TELEX: 200797 NIXN UR  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1359 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Clostridium tetani  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..1356  
 ; US-08-280-228-3

Alignment Scores: 6.26e-36 Length: 1359  
 Pred. No.: 420.00 Matches: 131  
 Score: 44.20% Conservative: 90  
 Percent Similarity: 26.20% Mismatches: 169  
 Best Local Similarity: 17.57% Indels: 110  
 Query Match: 1 Gaps: 20  
 DB:



US-09-910-186a-10 (1-450) x US-08-280-238-3 (1-1359)

```
Qy 6 AsnIlePheSerThrAsnAsnSerLeuLeuLysAspIleLeuAsnGluThrPheAsn 25
Db 7 AACCTGATTGTTGGTCGCAACAGAGACATCGATGTTATC 51
Qy 26 AsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThr 45
Db 52 ---CTGAAAAGTCTACCATCTCTGAACCTGGACATCAACAGCATATTATCTCCGACATC 108
Qy 46 SerGlyThrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePhePro 65
Db 109 TCTGTTTCAACTCTCTGTTATCACAATCCAGATCTCAATGTGGCG 159
Qy 66 PheAspPheLysLeuGlySerSerGlyGlu 78
Db 160 -----GGCATCAACGGCAAGCTATCCACCTGCTTAACACGAATCTTCT 204
Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98
Db 205 GAAGTTATCGTGCACAAGGCCATGACATCAATACACATGTTCAACACATTTCAAC 264
Qy 99 IleSerPheTrpIleArgIleAsnLys---TrpValSerAsnLeuPro 113
Db 265 GTTAGCTTCTGGTGGCGTTCGAAAGTTCTGCTTCCACCTGGACAGTACGCGACT 324
Qy 114 ---GlyThrIleLeuAspSerValLysAsnAsn 127
Db 325 AACGAGTACTCCATCATCAGCTCTATGAAGAAACATCCCTGCTCCATCGGCTCTGGTTGG 384
Qy 128 SerIleGlyIleLeuSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGlu 147
Db 385 TCTGTTCCCTGAAGGTAAACACTGATCTGGCTGACTGAACTCCGCGCGGCGAAGTT 444
Qy 148 GlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr-----AsnLys 165
Db 445 CGTCAGATCACTTTCCCG---GACCTCGCGACAAAGTTCAACGCTACCTGGTGAACAA 501
Qy 166 TrpPhePheValThrValThrAsnAsnMetMetClyAsnMetLysIleTyrIleAsnGly 185
Db 502 TGGGTTTCACTACATCAACAGATCGGTCTGCTGCTTCTAACCTGTACATCAACCGC 561
Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205
Db 562 GTTCGTATGGCTCCGCTGAATAACTACTGGTCTGGCGCTATCCGTGAGGACACAAACATC 621
Qy 206 ThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsnIle 225
Db 622 ACTCTTAAGCTGGACCGT-----TCCACACAAACAAACCCAG 657
Qy 226 AsnMetTrpIleArgAspPheThrIlePheAlaLysGluLeuAspGlyLysAspIleAsn 245
Db 658 TAGGTATCCATCGACAAGTTCGGTATCTTCTGCAAGCACTGAACCCGGAAGAGATCGAA 717
Qy 246 IleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAsnAspLeu 265
Db 718 AAACGTATACAGCTACCTGCTCTATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 777
Qy 266 ArgTyrAsnLysGluTyrTyrMetValAsnIle 276
Db 778 CGTTACGACACCGGATATACCTGATCCGGTACTCTAGCTCTAAGACGTTACGCTG 837
Qy 277 -----AspTyrLeu-----AsnArgTyrMetTyrAlaAsnSerArgGln 289
Db 838 AAAAATCATCACTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 897
Qy 290 IleValPheAsnThrArgAsnAsnAsnAspPheAsnGluGlyTyrIleIleIle 309
Db 898 AACATCTACTACCGAGCTGTACAAC-----GGCTGAAATTCATCATC 942
Qy 310 LysArgIleArgGlyAsnThrAsn---AspThrArgValArgGlyGlyAspIleLeuTyr 328
Db 943 AAACGCTACCTCCGAAACAGCAATCGATTCTTCTGTTAAATCTGTTGCTACTTCATCAA 1002
```

Search completed: November 7, 2002, 19:11:48

Job time : 77 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 18:38:12 ; Search time 56 Seconds

(without alignments)  
2850.955 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTIFNFTSTNNSLLKDII.....NYASLLESTHWGFPVYSE 450

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09910186/runat\_04112002.111619\_20956/app\_query.fasta\_1.647  
-DB=Published\_Applications\_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BIOSUM62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09910186.ecgn\_1\_1\_18.@runat\_04112002.111619.20956  
-NCPJ=6 -ICPU=3 -NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2381	99.6	1371	10	US-09-350-756-3 Sequence 3, Appl1
2	2364	98.9	3876	12	US-10-051-952-4 Sequence 4, Appl1
3	863.5	36.1	3831	12	US-10-051-952-5 Sequence 5, Appl1
4	588.5	24.6	3894	12	US-10-051-952-9 Sequence 9, Appl1

5	586.5	24.5	4835	10	US-09-288-326-10 Sequence 10, Appl1
6	585.5	24.5	1400	10	US-09-350-756-4 Sequence 4, Appl1
7	582	24.3	3876	12	US-10-051-952-2 Sequence 2, Appl1
8	581.5	24.3	3759	12	US-10-051-952-6 Sequence 6, Appl1
9	581.5	24.3	3759	12	US-10-051-952-7 Sequence 7, Appl1
10	578.5	24.2	1347	10	US-09-350-756-6 Sequence 6, Appl1
11	568	23.8	2452	10	US-09-350-756-7 Sequence 7, Appl1
12	561	23.5	3876	12	US-10-051-952-3 Sequence 3, Appl1
13	557.5	23.3	1338	10	US-09-350-756-1 Sequence 1, Appl1
14	553.5	23.1	3891	12	US-10-051-952-1 Sequence 1, Appl1
15	550	23.0	1327	10	US-09-350-756-9 Sequence 9, Appl1
16	546	22.8	1293	8	US-08-981-087A-5 Sequence 5, Appl1
17	546	22.8	1313	8	US-08-981-087A-6 Sequence 6, Appl1
18	541	22.6	1341	10	US-09-350-756-2 Sequence 2, Appl1
19	529	22.1	1317	10	US-09-350-756-5 Sequence 5, Appl1
20	516.5	21.6	3825	12	US-10-051-952-8 Sequence 8, Appl1
21	400.5	16.8	1987	10	US-09-350-756-8 Sequence 8, Appl1
22	393	16.4	702	10	US-09-350-756-10 Sequence 10, Appl1
23	168.5	7.0	621	10	US-09-350-756-11 Sequence 11, Appl1
24	145.5	6.1	6025	9	US-09-895-913A-11 Sequence 11, Appl1
25	142.5	6.0	1431	10	US-09-774-414-2 Sequence 2, Appl1
26	133	5.6	2796	10	US-09-815-242-4523 Sequence 4523, Ap
27	133	5.6	2799	10	US-09-815-242-8340 Sequence 8340, Ap
28	129	5.4	2523	10	US-09-815-242-4724 Sequence 4724, Ap
29	129	5.4	2526	10	US-09-815-242-8653 Sequence 8653, Ap
30	128	5.4	4310	10	US-09-740-274-5 Sequence 5, Appl1
31	121.5	5.1	15567	12	US-10-047-676A-3 Sequence 3, Appl1
32	121	5.1	1799	8	US-08-834-666A-19 Sequence 19, Appl1
33	120	5.0	640681	10	US-09-790-988-1 Sequence 1, Appl1
34	118	4.9	2979	10	US-09-815-242-4754 Sequence 4754, Ap
35	118	4.9	3009	10	US-09-815-242-8801 Sequence 8801, Ap
36	118	4.9	2376	10	US-09-815-242-9060 Sequence 9060, Ap
37	116.5	4.9	2376	10	US-09-880-107-3276 Sequence 3276, Ap
38	116.5	4.9	2988	10	US-09-815-242-4196 Sequence 4196, Ap
39	116.5	4.9	3030	10	US-09-815-242-8043 Sequence 8043, Ap
40	116.5	4.9	4389	12	US-10-044-090-611 Sequence 611, App
41	116.5	4.9	4531	10	US-09-925-300-342 Sequence 342, App
42	113.5	4.7	4594	10	US-09-070-927A-230 Sequence 230, App
43	111.5	4.7	1866	10	US-09-856-247A-1 Sequence 1, Appl1
44	111.5	4.7	2367	10	US-09-995-587A-2 Sequence 2, Appl1
45	111.5	4.7	2394	10	US-09-995-587A-3 Sequence 3, Appl1

ALIGNMENTS

RESULT 1  
US-09-350-756-3  
; Sequence 3, Application US/09350756  
; Patent No. US20020034521A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP RIID 98-21  
; CURRENT APPLICATION NUMBER: US/09/350,756  
; CURRENT FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092,416  
; EARLIER FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 3  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
US-09-350-756-3

## Alignment Scores:

Pred. No.: 7,356-261 Length: 1371  
 Score: 2361.00 Matches: 448  
 Percent Similarity: 99.78% Conservative: 1  
 Best Local Similarity: 99.56% Mismatches: 1  
 Query Match: 99.58% Indels: 0  
 DB: 10 Gaps: 0

US-09-910-186a-10 (1-450) x US-09-350-756-3 (1-1371)

Qy	1	MethrleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle	20
Db	10	ATCACCATCCCATCAACATCTTCTCTACACCAACAACCTCCCTGTGTGAAGACATCATC	69
Qy	21	AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn	40
Db	70	AACGAGTACTTCAACAACATCAACGACTCCAGATCTCTGCTCCCTCGAGAACCGTAAAGAC	129
Qy	41	ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu	60
Db	130	ACCTTGGTCGACACCTCCGGTTACACGCCGAGGTCTCCGAGGAGGTGACCTCCAGCTG	189
Qy	61	AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal	80
Db	190	AACCCAAATCTTCCCATTCGACTTCAAGCTGGGTTCTCCGGTGAGGACAGAGTAAAGTC	249
Qy	81	IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer	100
Db	250	ATCGTCACCCAGAGACGAGAACATCGTACAACTCCATGTACGAGTCCCTTCCATCTCC	309
Qy	101	PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleLeuAspSer	120
Db	310	TTCTGGATCAAGATCAACAAGTGGGTCTCCAACTTCCGAGTTACACCATCATCGACTCC	369
Qy	121	ValLysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeu	140
Db	370	GTCAGAACAACTCCGGTGTGTCATCGGTATCATCTCCACCTTCTCGGTCTTCAACCTG	429
Qy	141	LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla	160
Db	430	AAGCAGACGAGGACTCCGACGATCCATCACTTCTTCCAGCATCTCCACACACGCT	489
Qy	161	ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys	180
Db	490	CTGTGTACCAACAGTGGTTCTTCTGTCACCGTCACCAACAACATGATGGTAAACATGA	549
Qy	181	IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn	200
Db	550	ATCTACATCAACGGTAAGCTGATCGACACCATCAAGGTCAAGGATGACCGGTATCAAC	609
Qy	201	PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer	220
Db	610	TTCTCCAGACCATCACTTCGAGATCAACAAGATCCACACACCGGTCTGATCACCTCC	669
Qy	221	AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp	240
Db	670	GACTCCGACAATATCAACATGTGGATCGGTGACTTCTACATCTTCGCCAAGGAGTTGGAC	729
Qy	241	GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyr	260
Db	730	GTAAGGACATCAACUCTGTTCACTCTTGGAGTACACCAACGTCGTCAGAGCTAC	789
Qy	261	TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetValAsnIleAspTyrLeuAsn	280
Db	790	TGGGTACGACGCTGAGATACAAAGAGGTACTPACATGTGTCACATCGACTACTTGAAC	849
Qy	281	ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp	300
Db	850	AGATACATGTACGCCAACTCCAGACAGATGCTCTTCAACACCCAGACGCTACACACGAC	909
Qy	301	PheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArg	320
Db	910	TTCAACGAGGGTTACAAGATCATCAAGCGTTACAGAGGTATCAGAGGTAAACACCAACGAC	969

Qy 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340  
 Db 970 GTCAGAGGTGGTGACATCTCTGTACTTTCGACATGACTATCAACAAGGCGCTACACCTG 1029  
 Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360  
 Db 1030 TTCATGAAGAACAGAGACCATGTACGCCGACCAACCATCTCCACGAGGACATCTACGCCATC 1089  
 Qy 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380  
 Db 1090 GGTCTGGTGAGCAGACCAAGGACATCAACAGACATCATCTTCCAGATCCAGCCATG 1149  
 Qy 381 AsnAsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400  
 Db 1150 AACAACTTACTTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAGAACATC 1209  
 Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHis 420  
 Db 1210 TCCGGTATCTGTTCATCGGTACTACAGATTCGGTCTGGTGGTACTGGTACAGACAC 1269  
 Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuSerThrSer 440  
 Db 1270 AACTACTTGTTCCTCACTGTCAAGCAGGTTAACTACGCTCTCTGTGGAGTCCACTTCC 1329  
 Qy 441 ThrHisTrpGlyPheValProValSerGlu 450  
 Db 1330 ACCACTGGGGATTCGTCCTCCAGTCTCCGAG 1359

RESULT 2  
 US-10-051-952-4  
 ; Sequence 4, Application US/10051952  
 ; Patent No. US20020107199A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Walker, Patricia  
 ; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
 ; FILE REFERENCE: 2933CIP  
 ; CURRENT APPLICATION NUMBER: US/10/051,952  
 ; PRIOR FILING DATE: 2002-01-17  
 ; PRIOR APPLICATION NUMBER: 09/730,237  
 ; PRIOR FILING DATE: 2000-12-05  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: Patent in Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 3876  
 ; TYPE: DNA  
 ; ORGANISM: botulinum toxin  
 US-10-051-952-4

Alignment Scores:  
 Pred. No.: 2,8e-258 Length: 3876  
 Score: 2364.00 Matches: 447  
 Percent Similarity: 99.55% Conservative: 0  
 Best Local Similarity: 99.55% Mismatches: 2  
 Query Match: 98.87% Indels: 0  
 DB: 12 Gaps: 0

US-09-910-186a-10 (1-450) x US-10-051-952-4 (1-3876)

Qy	2	ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn	21
Db	2527	ACAATACCCITTAATATTTTTCATATACATAAATAATCTTTATTAAGATATAATAAT	2586
Qy	22	GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr	41
Db	2587	GAATATTCATAATAATATTAATGATTCAANAATTTTGAGCTACAAAACAGAAAAATACT	2646
Qy	42	LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlyAspValGlnLeuAsn	61
Db	2647	TTAGTGGTACATCGGATATATATGCAAGAGTGAAGTCAAGAGCGCATGTTCACTTAAT	2706
Qy	62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle	81

```
Db 2707 CCAATATTTCATTGCTTTAAATTAGGTAGTTCAGGGGAGGATAGAGTAAAGTTATA 2766
Qy 82 ValThrGlnAsnGluAsnIleValTyArgSerMetTyArgSerPheSerIleSerPhe 101
Db 2767 GTAACCCAGAAATGAATATGTTATATATATATATATATATATATATATATATAT 2826
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleLeuAspSerVal 121
Db 2827 TGGATTAGATAAATAAATGGTAAAGTAAATTTACCTGGATATATATATATATATAT 2886
Qy 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 2887 AAAATAAATCACTGGGTGGAGTATAGGTATATATATATATATATATATATATAT 2946
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyArgSerAsnAsnAlaPro 161
Db 2947 CAAATGAAGATAGTGAACAAAGTAAATTTAGTTATCATATATCAAAATATGCTCT 3006
Qy 162 GlyTyArgAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3007 GGATAGAATAAATGGTTTTTTGTAAGTGTACTAACAAATATGATGGAAATATCAAGAT 3066
Qy 182 TyRileAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3067 TATATAATGGAAATTAATAGATATCTAAAGTTAAAGACTAACTGGAATTAATTT 3126
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3127 ACCAAATATATAAATTTGAATATAATATAATCCAGATACCGGTTTGATTAATCA 3186
Qy 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyRilePheAlaLysGluLeuAspGly 241
Db 3187 TCTGTATACATCAATATGCTGGATAGAGATTTTATATATTTGCTAAAGAAATAGAT 3246
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValValLysAspTyTrp 261
Db 3247 AAGATATATATATATATTAATAGTTCGAATATATACTAAATGTTGTAAGAAAT 3306
Qy 262 GlyAsnAspLeuArgTyArgLysGlyTyThrMetValAsnIleAspTyRileAsnArg 281
Db 3307 GGAATGATTTAAGATATATAAAGAAATTTATATGTTTAAATATAGATTTATTAAT 3366
Qy 282 TyrMetTyRalaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301
Db 3367 TATATGATGCAACTACGACAAATTTGTTTAAATACAGTACAAATATAATAGCTTC 3426
Qy 302 AsnGluGlyTyRileIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
Db 3427 AATGAAGGATATAAATTAATAAAGAAATCAGAGAAATACAAATGATAGTAGTA 3486
Qy 322 ArgGlyAspIleLeuTyRArgPheAspMetThrIleAsnAsnLysAlaTyArgLeuPhe 341
Db 3487 CGAGGAGGAGATATTTATATTTGATATGACAATTAATAACAAGCATATAATTTGTT 3546
Qy 342 MetLysAsnGluThrMetTyRalaAsnHisSerThrGluAspIleTyRAlaIleGly 361
Db 3547 ATGAAGAAATGAACATATGATGACATATATATATATATATATATATATATATAT 3606
Qy 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3607 TTAAGAGACAAACAAGGATATAATGATATATATATATATATATATATATATAT 3666
Qy 382 AsnThrTyTrpTyRAlaSerGlnIlePheLysSerAsnAsnGlyGluAsnIleSer 401
Db 3667 AATACATATATATAGCATCTCAAAATATTTAAATCAAAATTTTAAATGAGAAATAT 3726
Qy 402 GlyIleCysSerIleGlyThrTyArgPheArgLeuGlyGlyAspTyTrpTyArgHisAsn 421
Db 3727 GGAATATGTTCAATAGGATCTATCTGTTTACGTTGAGCTGGAGTGATGGTATAGACA 3786
Qy 422 TyrLeuValProThrValLysGlnGlyAsnTyRAlaSerLeuLeuGluSerThrSer 441
Db 3787 TATTTGGTCCTACTGTGAAGCAAGAAATATGTTTATATATATATATATATATATAT 3846
```

```
Qy 442 HisTrpGlyPheValProValSerGlu 450
Db 3847 CATGGGGGTTTGTACCTGTAGTGAA 3873

RESULT 3
US-10-051-952-5
; Sequence 5, Application US/10051952
; Patent No. US20020107199A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 3831
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-5

Alignment Scores:
Pred. No.: 4,71e-88 Length: 3831
Score: 863.50 Matches: 196
Percent Similarity: 61.20% Conservative: 90
Best Local Similarity: 43.46% Mismatches: 148
Query Match: 36.11% Indels: 27
DB: 11 Gaps: 11

US-09-910-186a-10 (1-450) x US-10-051-952-5 (1-3831)
Qy 2 ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2515 ACAATGCCTTTTAAATATATTTTTCATATACTAATAATATCTTTTATAAAGATATAAT 2574
Qy 22 GluTyRAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2575 GAATATTTCAATAGTATTAATGATTCACAAATTTTGAGCTTACAAACACAAATGCT 2634
Qy 42 LeuValAspThrSerGlyTyRAsnAlaGluValSerGluGlyAspValGlnLeuAsn 61
Db 2635 TTAGTGGATACATCAGGATATAATGCAGAGTGAGGCTAGGAGATATATGTTCACTTA 2694
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2695 AGGATATATACAAATGACTTTAAATTAAGTAGTTCAGAGAT-----AAATTTATA 2745
Qy 82 ValThrGlnAsnGluAsnIleValTyArgSerMetTyRArgSerPheSerIleSerPhe 101
Db 2746 GTAAATTTAAATAATAATATTTTATATAGCGCTATTTATGAGAACTCTAGTGTAGTTT 2805
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGly---TyrThrIleIleAspSer 120
Db 2806 TGGATTAAGATATCTTAAAGATTTAACTAAATTTCTCATATGAATATACAAATAAAGCT 2865
Qy 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
Db 2865 ATAGACAAATTTCTGGTGGAAATTTATGATAGGAATGCAATATAGATGGATTTTA 2925
Qy 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyArgSerAsnAsnAla 160
Db 2926 CAAAGATGTTAATAGAAAGTATAAAGTTTAAATTTTATATAGTGAATCATTAAGTCAT 2985
Qy 161 ProGlyTyR---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet 179
Db 2986 ACAGGATATACAAATPAATAGTTTTCCTTACTACTAATATATATATATATATATG 3045
Qy 180 LysIleTyRileAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199
Db 180 LysIleTyRileAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIle 199
```

Db 3046 AAACCTTTATATATGAGAAATTAAGCAGAGTCAAAATTAAGAGATTAGAGTGT 3105  
Qy 200 AsnPheserlythrlethrPheGluilleAsnLysIleProaspThrClyLeullethr 219  
Db 3106 AAGTTAGATAAACCATAGTATTGGATGAGAGAGAAATATAGAT----- 3150  
Qy 220 SerAspSerAsnIleAsnMetTrpIleArgAspPheTyrllePheAlalysGluLeu 239  
Db 3151 -----GAGAAATCAGATCGCTTGGATTAGAGATTTTAAATATTTTCTTAAAGAAATTA 3201  
Qy 240 AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValLysasp 259  
Db 3202 AGTAATGAAATATTAATATGTTATATGAGGGACAAATATTAGAAATGTTATTAAAGAT 3261  
Qy 260 TyrTrpGlyAsnAspLeuArgTyAsnLysGluTyTrMetValAsnIleAspTyrlle 279  
Db 3262 TATTGGGAAATCTTGAAGTTTGATACAGATATATATATTATTAATGATATATATA 3321  
Qy 280 AsnArgTyMetTyrlleAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsn 299  
Db 3322 CATAGGTATATTCACCTGAAAGTAAATGTTACTTGTACTTGTGGTATCCAGATAGATCT 3381  
Qy 300 AspPheAsnGluGlyTyrlleLysIleIleLysArgIleArgGlyAsnThrAsnAspThr 319  
Db 3382 AAATATATATCTGAAATCTTATCTATTAAATCAGTATCAGTAAAGATCCTTATAGT 3441  
Qy 320 ArgValArgGlyLysPheIleLeuTyThrPheAspMetThrIleAsnAsnLysAlaTyAsn 339  
Db 3442 AGAATTTTAAATGAGGAGATATATAATTTCTCATATGTTATATATATAGTAAATATATG 3501  
Qy 340 LeuPheMetLysAsnGluThrMetTyrlleAsnHis-----SerThrGluasp 356  
Db 3502 ATATAAGAGATATCTGATACATATATATGACACACAGGAGGAGGTGTCACAAATGCT 3561  
Qy 357 IleTyrlleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIlelePheGln 376  
Db 3562 GTATATGCAATTAATACAGAGTAAATTTAGTAATATATGTTAGTATAGT---ATATTAGT 3618  
Qy 377 IleGlnProMetAsnAsnThrTyTrTyTrAlaSerGlnIlePheLysSerAsnPheAsn 396  
Db 3619 ATAAAAATATTTGATCTAAAAATTAATATTTGATGCAAAATTTTC---TCTAGTTTAGG 3675  
Qy 397 GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyArgPheArgLeuGlyGly 415  
Db 3676 ---GAAATACATACGCTCTAGCAGATATATATAAACTTGGAGATTTCTTTT----- 3726  
Qy 416 AspTrpTyArgHisAsnTyLeuValProThrValLysGlnGlyAsnTyrlleAsnLeu 435  
Db 3727 -----AAAAATGCATAGACGCCA---GTTCCAGTAATACTAATATGAACAAAA 3771

## RESULT 4

US-10-051-952-9  
; Sequence 9, Application US/10051952  
; Patent No. US20020107199A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Patricia  
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
; FILE REFERENCE: 2933CIP  
; CURRENT APPLICATION NUMBER: US/10/051,952  
; CURRENT FILING DATE: 2002-01-17  
; PRIOR APPLICATION NUMBER: 09/730,237  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 3894  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: primers used to introduce Stu I and EcoR I  
; OTHER INFORMATION: restriction sites into the 5' and 3' ends of the  
; OTHER INFORMATION: BONT/A-L chain gene fragment  
US-10-051-952-9

Alignment Scores:  
Pred. No.: 7.63e-57 Length: 3894  
Score: 588.50 Matches: 155  
Percent Similarity: 49.39% Conservative: 87  
Best Local Similarity: 31.63% Mismatches: 171  
Query Match: 24.61% Indels: 77  
Gaps: 17

US-09-910-186a-10 (1-450) x US-10-051-952-9 (1-3894)

Qy 2 ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
Db 2518 AGTATACCATTTGATCTTTCACATATATACCAAGCACAAATTTTAAACAAGTTTAAAT 2577  
Qy 22 GluTyThrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGluAsnArgLysAsnThr 41  
Db 2578 AATATATATTAGTATATAGTAAATGCTATTTTAAAGTTTAAAGTTATAGAGTGGCGGT 2637  
Qy 42 LeuValAspThrSerGlyTyAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
Db 2638 TTAATAGATTCATCTGGATATGTCGAACATATGAATGTCAGTTCAGATGTTATCTTTAT 2697  
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyLysAspArgGlyLysValle 81  
Db 2698 GATATAGGAATGCTCAATTTAAATTAATAATTCCT-----GAAATAGTAAATATAGC 2751  
Qy 82 ValThrGlnAsnGluAsnIleValTyAsnSerMetTyTrGluSerPheSerIleSerPhe 101  
Db 2752 GCATCATCAAGTAAATTCGTTGATATGATAGTATGTTTGATATTTTAGCATTAACITT 2811  
Qy 102 TrpIleArgIleAsnLysTrpValSerAsn-----LeuProGlyTyTr 115  
Db 2812 TGGTAAAGGACTCTCAATATAATAATAATATGATATACAACTTATCTTCAAAATGAGTAT 2871  
Qy 116 ThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPhe 135  
Db 2872 ACAATAATTTAGTTGATATAAAATGACTCAGGATGAAGATCTATTAAAGGGAATAGA 2931  
Qy 136 LeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrlle 155  
Db 2932 ATAATATGGACATTAATAGATGTTAATGCAAAATCTAAATATATTTTTCGAATATAGT 2991  
Qy 156 IleSerAsnAsnAlaProGlyTyTr---AsnLysTrpPhePheValThrValThrAsnAsn 174  
Db 2992 ATAAAAGATATATATCAGATTTATATAAATGATGTTTCCATAACTATTACTAATGAT 3051  
Qy 175 MetMetGlyAsnMetLysIleTyrlleAsnGlyLysLeuIleAspThrIleLysVallys 194  
Db 3052 AGATTAGTAAACGCAATATTTATATAATGGAAGTTTGAAAAAAGTGAATAAATTTTA 3111  
Qy 195 GluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProasp 214  
Db 3112 AACTTAGATAGATTAATCTAGTAATGATATAGACTTCARATTAATTAATTAATGTCAGAT 3171  
Qy 215 ThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrlle 234  
Db 3172 ACTACTAAATTT-----GTTTGGATTAAGGATTTTAAATATT 3207  
Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThr 254  
Db 3208 TTTGGTAGAGAAATTAATGCTACAGAAATGCTATCTCATATATTTGGATTCATCATCTACA 3267  
Qy 255 AsnValValLysAspTyTrPglYAsnAspLeuArgTyTrAsnLysGluTyTrMetVal 274  
Db 3268 AATACTTTAAAGATTTTGGGGGAATCCCTTTAAGATAGGATACACAAATACTATCTGTT 3327  
Qy 275 AsnIleAspTyTrLeuAsnArgTyTrMetTyrlleAsnSerArgGlnIleValPheAsnThr 294

```
Db 3328 AATCAAGGATGCGAAATATCATATATAAGATTTTAAAGCTTCTATGGGGAACT 3387
QY 295 AAGATGAsnAsnAspPheAsnGlu-----Gly 304
Db 3388 GCACACGACAAAC--TTTAATAATGCAGCAATAAATATCAAAATTTATATCTTGGT 3444
QY 305 TyrlYsIlleIlleIleYsArgIle-----ArgGlyAsnThrAsnAspThrArgVal 321
Db 3445 TTACGATTTATATAAAAAAGCATCAAAATCTCGGAATATAATAATGATATAATAGTC 3504
QY 322 ArgGlyGlyAspIleLeuTyrlPheAspMetThrIleAsnAsnLysAlaTyrlAsnLeuPhe 341
Db 3505 ACAGAAGGAGATATATATATCTTAATTT----- 3534
QY 342 MetLysAsnGluThrMetTyrlAlaAspAsnHisSerThrGluAsp-----IleTyrlAla 359
Db 3535 -----GATAAATATTTCTGATGAATCTTAGAGAGATATATGTT 3570
QY 360 IleGlyLeuArgGluInThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPro 379
Db 3571 TTG-----GTCAATCTTAAGAAATTCAAACTCAATATTT-----TTAGCACCC 3615
QY 380 MetAsnAsn-----ThrTyrlTyrlAlaSerGlnIle-----PheLysSer 393
Db 3616 ATAATGATGATCCTACGCTTCTATGATCTACTACAAATAAAAAAATATATGAAAAACA 3675
QY 394 AsnPheAsnGlyGluAsnIleSer-----GlyIleCysSerIle 406
Db 3676 ACATATAITGTCAGATACTTTGGCAAAAGATACTAAACATTTGGCGTGTGGNATT 3735
QY 407 GlyThrTrpArgPheArgLeuGlyAspTrpTyrlArgHis-----AsnTyrlLeuValPro 425
Db 3736 GGTAAATTTGTTAAAGATATGATATCTTTGGGATACCTATGATGATAATTTATTTT 3789
QY 426 ThrValLysGlnGlyAsnTyrlAlaSerLeuLeuGluSerThrSer----- 440
Db 3790 TGCATAGTCAGTGTACTCTCAGAGAATATCTCGAAATATAATAAATTAAGTTGGGA 3849
QY 441 ThrHisTrpGlyPheValProValSerGlu 450
Db 3850 TGTAATGGCAATTCATTCCTCCGCTGGATGAA 3879
```

## RESULT 5

```
US-09-288-326-10
; Sequence 10, Application US/09288326
; Patent No. US20010018049A1
; GENERAL INFORMATION:
; APPLICANT: Kei Roger Aoki
; APPLICANT: George Sachs
; TITLE OF INVENTION: Method and Compositions for the
; FILE OF INVENTION: Treatment of Pancreatitis
; FILE REFERENCE: 17282
; CURRENT APPLICATION NUMBER: US/09/288.326
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 4835
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-09-288-326-10
```

## Alignment Scores:

Pred. No.:	1,76e-56	Length:	4835
Score:	586.50	Matches:	149
Percent Similarity:	48.94%	Conservative:	82
Best Local Similarity:	31.57%	Mismatches:	186
Query Match:	24.53%	Indels:	55
DB:	10	Gaps:	11

US-09-910-186a-10 (1-450) x US-09-288-326-10 (1-4835)

```
QY 3 IleProPheAsnIlePheSerTyrlThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
Db 2902 ATACCTTTTCAGCTTTCCAAATACGTAGATAAATCAAGAGATTATTAATCTACATTACTGAA 2961
QY 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
Db 2962 TATATTAAAGAAATATTATTAATCTCTATATTAATTTAAAGATGAAGAAATCAATTTA 3021
QY 43 ValAspThrSerGlyTyrlAsnAlaGluValSerClucluglyAspValGlnLeuAsnPro 62
Db 3022 ATAGACTTATCTAGTATGTCATCAAAATAAATATTGGTAGTAAGTAATTTTGATCCA 3081
QY 63 IlePheProPheAsp-----PheLysLeuGlySerSerGlyGluAspArgGlyLys 79
Db 3082 ATAGATAAAATCAAAATCAATTTATTTAGAAAGTAGT-----AAA 3126
QY 80 ValIleValThrGlnAsnGluAsnIleValTyrlAsnSerMetTyrlGluSerPheSerIle 99
Db 3127 ATTTGGGTAATTTTAAAAAATGCTATTGTTATATAAATAGTATGATAAAATTTTAGTACT 3186
QY 100 SerPheTrpIleArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrlThr 116
Db 3187 AGCTTTGGATGAAGATTTCTTAAGTATTTAACAGTATAAGTCTAAATAATGAATATACA 3246
QY 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
Db 3247 ATAATAAATTTCTATGAAAAATAATTCAGATGGAAGTATCACTTAATTTATGTTGAAATA 3306
QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrlAspIle 156
Db 3307 ATCTGGACTTTACAGGATCTCAGGAATAAACAAGAGTAGTTTTTAAATACATCACTCAA 3366
QY 157 SerAsnAsnAlaProGlyTyrl---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
Db 3367 ATGATTATATATATCATGATATATAAACAAGATGGATTTTGTACTATCACTCAATAATAGA 3426
QY 176 MetGlyAsnMetLysIleTyrlIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
Db 3427 TTAATAACTCTAAATTTATATAAGTGAAGATTATAGATCAAAACCAATTTCAAAT 3486
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 3487 TTAGGTAATATTCATGCTAGTAAATAATATATGTTTAAATAGATGTTGTAGAGATACA 3546
QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetIrlPheArgPheTyrlIlePhe 235
Db 3547 -----CATAGATATATTGGATAAAATATTTTAATCTTTTT 3582
QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrlThrAsn 255
Db 3583 GATAAGGAATTAATGAAAAAGAAATCAAGATTTATATGATAATCAATCAAAATTCAGGT 3642
QY 256 ValValLysAspTyrlTrpGlyAsnAspLeuArgTyrlAsnLysGluTyrlMetValAsn 275
Db 3643 ATTTTAAAGACCTTTGGGTGATTTTACAATATGATAAACCATCTATATGTTAAT 3702
QY 276 IleAspTyrlLeuAsnArgTyrlMetTyrlAlaAsnSer----- 287
Db 3703 TTATATGATCCAAATAAATATGTCGATGTAATAATAGTAGTATTAGAGGTTATATGTAT 3762
QY 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsn 302
Db 3763 CTAAAGGGCCTAGAGGTAGCGTAAATGACTACAAACATTTTATTAAATCAAGTTCTGAT 3822
QY 303 GluGlyTyrlLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArg 322
Db 3823 AGGGGACAAAAATTTATTAATAAAAAATATCTCTCGAAATAAAGATAATATTGTTAGA 3882
QY 323 GlyGlyAspIleLeuTyrlPheAspMetThrIleAsnAsnLysAlaTyrlAsnLeuPheMet 342
Db 3883 AATAATGATCGTGTATATATTAATGCTAGTAGTTAAATAAAGAAATATAGCTTA----- 3936
QY 343 LysAsnGluThrMetTyrlAlaAspAsnHisSerThrGluAspIleTyrlAlaIleGlyLeu 362
```

```

Db 3937 -----GCTACTAATGCACACAGCGAGGTAGAAAAATACTAAGT 3978
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGluIleGlnProMetAsnAsn 382
Db 3979 GCATTAGAAATACCTGATGAGAAATCTAAGTCAAGTAGTAGTAAGTCAAAAAAT 4038
QY 383 ThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
Db 4039 GATCAAGGAATAACAATAATGC--AAATGAATTTACAAGATAATAATGGGAATGAT 4095
QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyClyAspTrpTyrArgHisAsnTyr 422
Db 4096 ATAGCGTTTATAGATTTCATGAGTTT-----AATAAT 4128
QY 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 4129 ATACCTAACTAGTAGCAAGTAATGGTATATAATAGACAAATAGAAAGATCTAGTAGGACT 4188
QY 441 -----ThrHisTrpGlyPheValProValSerGlu 450
Db 4189 TTGGGTGCTCATGGGAATTTATTCCTGTAGATGAT 4224

```

## RESULT 6

```

US-09-350-756-4
; Sequence 4, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 4
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-4

```

## Alignment Scores:

```

Pred. No.: 3,81e-57 Length: 1400
Score: 585.50 Matches: 152
Percent Similarity: 51.72% Conservative: 89
Best Local Similarity: 32.62% Mismatches: 160
Query Match: 65 Indels: 65
DB: 10 Gaps: 17

```

US-09-910-186a-10 (1-450) x US-09-350-756-4 (1-1400)

```

QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 64 ATATTCCTTTTAACTCTCTCTTATACAGATGATATAAATTAAATTCCTACTTCAAC 123
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 124 AAGTCTCTCAAGAGAATTAAGTCTCTCCGTTTAAACATGAGATACAAAGATGATAAA 183
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 184 TAGCTCGACACTTCGGTTAGCATCTCAATATCAACATTAACTGCTGACGTGTAAGATAC 243

```

```

QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 244 CCAACTAACAACAAACCAATTC-----GCTATCTACAACGCAAGCTTACTGAGGTGAAC 297
QY 82 ValThrGluAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 298 ATCTCTCAAAACGACATACATATCTACGACAACAGTAGTACAAGAACTCTCTATTTCTTC 357
QY 102 TrpIleArgIle-----AsnLysTrpValSerAsnLeuProGlyTyrThrIle 117
Db 358 TGGGTCAAGGATTCCTTAACAGCAACAAGATCGTCAACGTTTAAACAAGAGTACACATCT 417
QY 118 IleAspSerValLys---AsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeu 136
Db 418 ATCAACTGTTATGAGACACAACACTCCGTTGGAAGTCTCTCTTAACCAACAGAGATC 477
QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
Db 478 ATTGGACCTTGAACAGACACGCGAGGTATTAAACCAAAAGTTAGCATTCAACTACGCTAAC 537
QY 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnMet 175
Db 538 GCAACGCTATTCTGACTACATCAACAAGTGGATTTCTGCTACTATCACTACACAGACA 597
QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
Db 598 TTAGTGACTCTAAGCTTTTACATTAACGGTTAACTTAATCGACCAAAAGTCCCATTTAAAC 657
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 658 TTAGTAAACATTCACGTTCTGACAACATCTTATTCAAGATCGTTAACTGCGATTTACACC 717
QY 216 GlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235
Db 718 AGATACATT-----GGCATTAGATACITCAACATTTTC 750
QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIlePheAsnSerLeuGlnTyrThrAsn 255
Db 751 GACAAGGATTTAGACAGACGAGATTCAACTTATACAGCAACGAACTAACACCAAT 810
QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
Db 811 ATTTTGAGGACTTCGCGGTACTACTGCTTTACGACAAGGAATACTACTATTATAAC 870
QY 276 IleAspTyrLeuAsnArgTyrMet-----TyrAlaAsnSerArgGlnIleValPheAsn 293
Db 871 GTGTTAAAGCCAAACAACTTCATTGATAGGAGAAAGGATTCCTACTTTAAGCATTAACAAC 930
QY 294 ThrArgArgAsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIleLys 310
Db 931 ATCAGAAGCCTATTCTTTAGCTAACAGATTATCTCTGGTATCAAGGTTAAGATCCAA 990
QY 311 ArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyAspIleLeuTyrPhe 329
Db 991 AGAGTTAAACAACCTCTCTACTAACGATAACCTTTGTAGAAAGAACGATCAGGTCTATAT 1050
QY 330 AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla 349
Db 1051 AACTTCGCTGCTAGCAAGACT-----CACTTATTC-----CCATTATATGCT 1092
QY 350 AspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIle 369
Db 1093 GATACCGCTACCAACAAC-----AAGGAGAAGACCATCAAGATC 1131
QY 370 AsnAsp-----AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrAla 387
Db 1132 TCCCTCCTCTGGCAACAGATTTAAACCAATCGGTCTGTTATGAATCCCTCGGTACAACTGT 1191
QY 388 SerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGly 407
Db 1192 ACCATGAACCTTTAAATAAT---AATGGAATAAT-----ATTGGG 1230
QY 408 ThrTyrArgPheArg-----LeuGlyClyAspTrpTyrArgHisAsnTyrLeu 423

```

```

Db 1231 TGTTAGTTTCACAGGAGATAGTCTAGTTGCTAGTACTGGTAT----- 1275
Qy 424 ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHis--- 442
Db 1276 -----TATACCCACATGAGAGATCACACCACACAGCATGGA 1311
Qy 443 -----TrpGlyPheVal 446
Db 1312 TGTTTTGGAACTTTATT 1329

RESULT 7
US-10-051-952-2
; Sequence 2, Application US/10051952
; Patent No. US20020107199A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; PRIOR FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3876
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-2

Alignment Scores:
Pred. No.: 4,14e-56 Length: 3876
Score: 582.00 Matches: 161
Percent Similarity: 51.33% Conservative: 89
Best Local Similarity: 33.06% Mismatches: 163
Query Match: 24.34% Indels: 74
Db: 12 Gaps: 20

US-09-910-186a-10 (1-450) x US-10-051-952-2 (1-3876)
Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22
Db 2506 ATACCTTTGATCTTTCACGTATCTTAATATGCAATATCTATAAATAATTAATAA 2565
Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42
Db 2566 TATAATACGGAATTTTAAATAATATTATCTTAATTTAAGATATAGAGATAATAATTA 2625
Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62
Db 2626 ATAGATTATCAGGATATGGCAAGAGTAGATATATGCGGTCAAGCTTAATGAT 2685
Qy 63 IlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleVal 82
Db 2686 -----AAAATCAATTTAAATTAAGTATGTCACAGAT-----AGTAAGATTAGATC 2733
Qy 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102
Db 2734 ACTCAAAATCAGATATATATTAATAGATATGTCCTGATTTTACGGTAGCTTTTGG 2793
Qy 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116
Db 2794 ATAAGATACCTAAATAGGAATGATGATATACAAATATATATTCATATGAATATACG 2853
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
Db 2854 ATAATTAATGTATGAAATAAATAATTCAGGTGGAATAATCTATTAGGGGTAAATAGATA 2913
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
Db 2914 ATATGACCTTAATGTATATAAATGGAAAAACCAATCAGTATTTTGTGATATACATA 2973

```

## RESULT 8

US-10-051-952-6

; Sequence 6, Application US/10051952

; Patent No. US20020107199A1

; GENERAL INFORMATION:

```

Qy 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
Db 2974 AGAAGAGATATATACAGAGTATATAAATAGATGTTTTTTGTAACTATTACTATAAT--- 3030
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGlu 195
Db 3031 TTGGATAATGCTAAATTTATATTAATGACCGTTAGATCAATCAATATGATATTAAGAT 3090
Qy 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysLysIleProAspThr 215
Db 3091 ATAGGAAGATTTATTTGTTAATGTTGAATTAACATTTAAATTTAGAT----- 3135
Qy 216 GlyLeuIleThrSerAspSerAspAsnIleAsn---MetTrpIleArgAspPheTyrIle 234
Db 3136 -----GGTGAATGATAGACACAAATTTATTTGGATGAATATTTTAGTATT 3183
Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThr 254
Db 3184 TTTAATACGCAATTAATCAATCAATATATAAAGAGATATATAAATTCATCATATAGC 3243
Qy 255 AsnValLysLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTrpMetVal 274
Db 3244 GAATAGTTAAAGATTTTGGGAATCCCTTTAATGTATAATAAAGAAATATTATATGTTT 3303
Qy 275 AsnIleAspTyrLeuAsnArgTyrMetTyr-----AlaAsnSerArgGlnIle 290
Db 3304 AATCGGGGAATAAAATTCATATATTAAACTAGTAGAAGATTCATCTGAGTGAATA 3363
Qy 291 ValPheAsnThrArgArgAsnAsnAsn-----AspPheAsnGlu 303
Db 3364 TTAATACGTAGCAAAATATATACAGATTCCAATATATAAATATAGAAATTTATATATT 3423
Qy 304 GlyTyrLysIleIleIleLysArgIleArgGly-----AsnThrAsnAspThrArgVal 321
Db 3424 GGAGAAAAATTTATTATAAGAGAGAGAGTCAATTCATCAATCTATAAATCATGATATAGTT 3483
Qy 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLys----- 336
Db 3484 AGAAAAGAAGATTATATACATCTAGATTTGGTACTTCACCATGAAGAGTGGAGATATAT 3543
Qy 337 AlaTyrAsnLeuPheMetLysAsnGlu-----ThrMetTyrAlaAspAsn 351
Db 3544 GCCTATAAATATTTTAAGGAACAGAGAAATAATGTTTTTATCTATTATAAGTGAT--- 3600
Qy 352 HisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAsp 371
Db 3601 ---TCTAATGAATTTTATAAGACTATATAAATAAAGAA----- 3636
Qy 372 AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrAlaSerGlnIlePhe 391
Db 3637 -----TATGATGACAGCCATCA-----TATAGTTGTCAGTTGCTTTT 3675
Qy 392 LysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrArgPhe 411
Db 3676 AAAAAA-----GATGAAGAAAGTACGATGATAGGATTTGATTCATTCCTTC 3729
Qy 412 ArgLeuGlyGly-----AspTyrTyrArgHisAsnTyrLeuVal----- 424
Db 3730 TAGGAATCTGAGCTTTTACGTAAAGATATAAAGATTTTGTATATAAGTAATAGTAG 3789
Qy 425 ---ProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrp 443
Db 3790 TTAAGAGAGGTAAAGAGAAACCATATATAAGTCAAAATTTG-----GGATGAATG 3840
Qy 444 GlyPheValProValSerGlu 450
Db 3841 CAGTTTATCTTAAAGATGAA 3861

```



```

; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 3753
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-6

```

```

Alignment Scores:
Pred. No.: 4.5e-56 Length: 3753
Score: 581.50 Matches: 153
Percent Similarity: 52.14% Conservative: 91
Best Local Similarity: 32.69% Mismatches: 153
Query Match: 24.32% Indels: 71
DB: 12 Gaps: 19

```

US-09-910-186A-10 (1-450) x US-10-051-952-6 (1-3753)

```

QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2464 AGTATCTCTTTAGCTTCTCTTTATACAGATGATTAATTTTATTCATATTTAAT 2523
QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2524 AAGTCTTTAAGAGATTAATAAGTAGTCTGTTTAAATATGATGATATAAAATGATAA 2583
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 2584 TAGTAGATACCTCAGGATATGATCAATATATAATATTAATGAGATGATATAAAATAT 2643
QY 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValle 81
Db 2644 CCAACTAATAAAATCAATTT-----GGAATATATAATGATAAACTTAGTGAAGTTAAT 2697
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 2698 ATATCTCAAAATGATTACATATATATGATAATAATAATTTTATGATTTAGTTT 2757
QY 102 TrpIleArgIle-----AsnLysTrpValSerAsnLeuProGlyTyrThrIle 117
Db 2758 TGGGTAGAATTCCTAACTATGATAATAAGATAGTAATGTTAATGATGATGATAATA 2817
QY 118 IleAspSerValLys---AsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136
Db 2818 ATAAATTTGATGAGGATAATAATTCAGGATGGAAGATATCTCTTAATCATATGAATA 2877
QY 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156
Db 2878 ATTTGGACATTCGAAGATAATTCAGGATTAATCAAAATACATTAATGATGTAAC 2937
QY 157 SerAsnAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175
Db 2938 GCAAAATGTTATTTCTGATTATATAATAAAGTGGATTTTCTGAATATACTAATGATAGA 2997
QY 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAsnIleAsnThrIleLysGlu 195
Db 2998 TTAGGAGATTTCAACTTATATTAAATGGAATTTAATATATAAAATCAATTTAAT 3057
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
Db 3058 TTAGTAAATATTCATGTTAGTGAATATATATTAAATAGTTAATTTAGTTATACA 3117
QY 216 GlyLeuIleThrSerAspSerAsnIleAsnMetTrpIleArgAspPheTyrIlePhe 235
Db 3118 AGATATATT-----GGTATTAGATATTTTAAATTTT 3150

```

```

QY 236 AlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGluTyrThrAsn 255
Db 3151 GATAAAGAATTAGATGAACACAGAAATTCAAATTTATATAACAATGAACCAATGCAAT 3210
QY 256 ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsn 275
Db 3211 ATTTTAAGGATTTTGGGAAATATTGCTTTATGACAAAGATAGTATTATTAAAT 3270
QY 276 Ile-----AspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleVal 291
Db 3271 GTGTTAAAAACCAATAAATTTATTATAGG-----AGAACAGATCTTACTTTAAGCAAT 3324
QY 292 PheAsnThrArgArgAsn-----AsnAsnAspPheAsnGluGlyTyrLysIleIle 308
Db 3325 AATATATATAGAGACACTATTCTTTTACCTATATAGATATATAGTGGATAAAAGTTAAA 3384
QY 309 IleLysArgIle---ArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeu 327
Db 3385 ATACAAAGAGTTAATAATATAGTACTACAGATAATCTGTTTAGAAAGATGATCAGGTA 3444
QY 328 TyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMet 347
Db 3445 TATATTAATTT---GTACGACGACAAACTCACTTACTT-----CCATTA 3486
QY 348 TyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLys 367
Db 3487 TAGCTGATACAGCTACCAAAAT-----AAAGAGAAACACATA 3525
QY 368 AspIleAsnAsp-----AsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyr 385
Db 3526 AAAATATCATCATCTGGCAATAGATTAATCAAGTAGTAGTATGATTAATTCAGTA----- 3579
QY 386 TyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSer 405
Db 3580 GGATGTCACATGAATTTTAAATAAAT---AATGGAATAAAT----- 3618
QY 406 IleGlyThrTyrArgPheArg-----LeuGlyGlyAspTrpTyrArgHisAsn 421
Db 3619 ATTTGGGTGTAGTTTCAAGCAGACTAGTGTGTGCTAGTACTTGGTAT----- 3669
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 3670 -----TATACACATATGAGAGATAATAACAAACAGC 3699
QY 442 His-----TrpGlyPheVal 446
Db 3700 AATGGATTTTGGGAACTTTTATT 3723

```

```

RESULT 9
US-10-051-952-7
; Sequence 7, Application US/10051952
; Patent No. US20020107199A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 3759
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-7

```

```

Alignment Scores:
Pred. No.: 4.52e-56 Length: 3759
Score: 581.50 Matches: 152
Percent Similarity: 51.72% Conservative: 89
Best Local Similarity: 32.62% Mismatches: 160

```

[illegible]

QY	330	AspMetThrIleAsnAsnLysAlaTyra	AsnLeuPheMetLysAsnGluThrMetTy	Ala 349
Db	3451	AAATTGTTAGCCAGCAAAACT	-----CACTTATTT-----	CCATTATATGCT 3492
QY	350	AspAsnHisSerThrGluAspIleTy	raIleGlyLeuArgGluInThrLysAsp	Ile 369
Db	3493	GATACAGCTACCACAAT	-----AAAGAAAAACAATAAAAAA	TA 3531
QY	370	AsnAsp-----AsnIleIlePheGln	IleGlnProMetAsnAsnThrTyTy	Ala 387
Db	3532	TCATCATCTGGCAATAGATT	TAATCAAGTAGTAGTATGAATTCAG	TAGGAATAAATTGT 3591
QY	388	SerGlnIlePheLysSerAsnPheAsn	GlyLeuAsnIleSerGlyIleCysSer	IleGly 407
Db	3592	ACAATGAATTTAAATAAT	--AATGAAATAAT--	-----ATTGGG 3630
QY	408	ThrTyraPheArg-----LeuGlyGly	AspTyrPtyrArgHisAsnTy	Leu 423
Db	3631	TTGTTAGGTTTCAAGCAGAT	ACTACTGAGTTGCTACTACTTGGTAT	----- 3675
QY	424	ValProThrValLysGlnGlyAsnTy	raLaserLeuLeuGluSerThrSerTh	His--- 442
Db	3676	-----TATACACATATGAG	ATCATCATACAAACAGCAATGGA	3711
QY	443	-----TrpGlyPheVal 446		
Db	3712	TGTTTTTGGAACTTTATT 3729		
RESULT 10				
US-09-350-756-6				
; Sequence 6, Application US/09350756				
; Patent No. US20020034521A1				
; GENERAL INFORMATION:				
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases				
; APPLICANT: John S. Lee				
; APPLICANT: Peter Pushko				
; APPLICANT: Michael D. Parker				
; APPLICANT: Jonathan F. Smith				
; APPLICANT: Mark T. Dertzbaugh				
; APPLICANT: Leonard Smith				
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine				
; FILE REFERENCE: 003/124/SAP RIID 98-21				
; CURRENT APPLICATION NUMBER: US/09/350,756				
; CURRENT FILING DATE: 1999-07-09				
; EARLIER APPLICATION NUMBER: US 60/092,416				
; EARLIER FILING DATE: 1998-07-10				
; NUMBER OF SEQ ID NOS: 11				
; SOFTWARE: Apple Macintosh Microsoft Word 6.0				
; SEQ ID NO 6				
; LENGTH: 1347				
; TYPE: DNA				
; ORGANISM: Artificial				
; FEATURE:				
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"				
US-09-350-756-6				
Alignment Scores:				
Pred. No.: 2,24e-56 Length: 1347				
Score: 578.50 Matches: 151				
Percent Similarity: 49.27% Conservative: 84				
Best Local Similarity: 31.66% Mismatches: 169				
Query Watch: 24.19% Indels: 73				
DB: 10 Gaps: 16				
US-09-910-186A-10 (1-450) x US-09-350-756-6 (1-1347)				

Qy	13	AsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAspSerLysIle	32
DB	7	GACACCATCTGATCCAGGCTTCAACATCATCTCCAACTATCTCTCCACGCCATC	66
Qy	33	LeuSerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluVal	52
pb	67	CTGTCCTCTGCTCAACCTGGTGGTGTCTGTATGCATCTCTCCGGTTCACGGACCACTTC	126

QY	53	SerGluGluGlyAspValGlnLeuAsnProIlePheProPheAspPhelysLeuGlySer	72
Db	127	AAGTGTCCGGTTCCTCAACGATCGGTGAAGGTTCAAGTCAAGCACAAC	186
QY	73	SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrsnSer	92
Db	187	TCC-----GAGAACCCTCAACATCACGCCGCCACCATCAAGTTCGGTCTACACACTCC	240
QY	93	MetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpValserAsn---	111
Db	241	ATGTCGACAACTTCCTCAACTTCCTGGTCCGTACCCCAAAGTACAAACAACACGAC	300
QY	112	-----LeuProGlyTyrThrIleIleAspSerValIysAsnSnSergly	126
Db	301	ATCCAGACCTTACCTGCAGAACGAGTAGTACACCATCATCTCTGTATCAAGAACGACTCCGGT	360
QY	127	TrpSerileGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluaspSer	146
Db	361	TGGAAGGTCTCCATCAAGGAAACCGTATCATCTGGACCCCTGATCGAGTCAACGCCAAG	420
QY	147	GluGlnSerIleasnPheserTyrAspIleSerAsnAsnAlaproGlyTyr---Asnlys	165
Db	421	TCCAAGTCCATCTTCCTCGAGTATCCATCAAGGACAAATCTCCGACTACATCAACAAAG	480
QY	166	TrpPhePheValThrValThrAsnMetMetGlyAsnMetLysIleTyrIleAsnGly	185
Db	481	TGGTTCCTCCATCACCATCACCAGACCGCTCTGGTGAACCCCAACATCTACATCAACGGT	540
QY	186	LysLeuIleasphrIleLysValLysGluLeuThrGlyIleasnPheserIystrille	205
Db	541	TCCTGTGAAGAGTCCGAGAAGATCTTGAACTGGACCGTATCAACTCTCCCAACGACATC	600
QY	206	ThrPheGluIleasnIylisIleProaspThrGlyLeuIleThrSerAspSerAspnille	225
Db	601	GACTTCAAGCTGATGTAAGTACCGACACCAACCAAGTTC-----	639
QY	226	AsnMetTrpIleargaspPheTyrIlePheAlaLysGluLeuaspGlyLysaspIleasn	245
Db	640	---GTCTGGATCAAGGACTTCACATCTCTGGTCTGGTGAAGCCACCGAGTCTCC	696
QY	246	IleLeuPheasnSerLeuGlnTyrThrAsnValIlyLysAspTyrTrpGlyAsnaspLeu	265
Db	697	TCCTGTACTGGATCCAGTCTCCACCACACCCCTGAAGGACTTCGGGGAACCCACTG	756
QY	266	ArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuasnArgTyrMetTyrAla	285
Db	757	CGTATCAGACACCCAGTACTACCTGTTCACACAGGATGATGAGAACATCTACATCAAGTAC	816
QY	286	AsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGlu-----	303
Db	817	TTCTCCAGGCTCCATGGGTGAGACCGCCCTCGTACCAAC---TTCACCAACGCCGCC	873
QY	304	-----GlyTyrLysIleIleIleLysargtle-----	312
Db	874	ATCAACTACCAAGAACCTGTACCTGGGTCTGCTTTCATCATCAAGAACGCCCTCCCACTCC	933
QY	313	ArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheaspMetThr	332
Db	934	CGTAACATCAACACAGCAACATCGTCCGTGGGTGACTACATCTACCTGACATC----	990
QY	333	IleAsnAsnLysAlatyAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHis	352
Db	991	-----GACACATC	999
QY	353	SerThrGluAspiletyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsn	372
Db	1000	TCCAGCAGTCC---TACCTGTGTACCTCTGGTCACTCCCAAGGAGATCCAGACCCAG	1056
QY	373	IleIlePheGlnIleGlnProMetAsnAsn-----ThrTyrTyrTyrAlaSerGlnIle	390
Db	1057	CTGTTC-----CTFGCCCCAAATCAACGACGACCCCTACCTTCACGACGCTCTCGCAGATC	1110

Qy 100 SerPheThrPheArgIleAsnLysTrpValSerAsnLeuPro-----GlyTyrThr 116  
Db 1401 AGCTTTGGATAAGAAATCTTAAGTATTTAACAGTATAAGCTAAATAATGAATATACA 1460  
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
Db 1461 ATAATAAATGTATGGAATAATCTCAGGATGGAAGATCATCTTAATATGGTGAATA 1520  
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
Db 1521 ATCTGGACTTACAGGATACCTCAGGAATAAACAAGAGTAGTTTAAATAC----- 1574  
Qy 157 SerAsnAlaProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMet 176  
Db 1575 AGTCA-AATGATTAAATATACACAGATGGATTGTGTAACATCATCTAATAATAGATTA 1633  
Qy 177 GlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeu 196  
Db 1634 AATAACTCTAAATTTATATAATGGAAGATTAAATAGATCAAAACCAATTTCAAAATTA 1693  
Qy 197 ThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGly 216  
Db 1694 GGTAATATTCATGCTAGTATATATAATCTTAAATAGATGGTTGTAGATACA--- 1750  
Qy 217 LeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAla 236  
Db 1751 -----CATAGATATATTTGGATAAAATATTTTAAATCTTTTGTAT 1789  
Qy 237 LysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVal 256  
Db 1790 AAGGAATTAATGAAAGAAATCAAGATTTATATGATTAATCAATCAATTCAGGTAT 1849  
Qy 257 ValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrMetValAsnIle 276  
Db 1850 TAAAGAACTTTGGGTGATTATTACAATATGATAAACCACTATATGTTAAATTA 1909  
Qy 277 AspTyrLeuAsnArgTyrMetTyrAlaAsnSer----- 287  
Db 1910 TATGATCCAAATAATATGCGATGAAATAATGTAGGTATATAGAGTTATATGATCATT 1969  
Qy 288 -----ArgGlnIleValPheAsnThrArg---ArgAsnAsnAsnAspPheAsnGlu 303  
Db 1970 AAAGGGCTAGAGGTAGGTAATCAGTACACAAACATTTATTTAAATCAAGTTGTATAGG 2029  
Qy 304 GlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGly 323  
Db 2030 GGCAGAAATTTATATAAAATAATGCTCTCGAAATAAGATATAATTTGTAGAAAT 2089  
Qy 324 GlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLys 343  
Db 2090 ATGATCGTGTATATATTAATGTAGTAGTAAATAAATAAGAAATATAGGTTA----- 2140  
Qy 344 AsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArg 363  
Db 2141 -----GCTACPAATGCATCAGCAGCGGTAGAAAATACTAAGTGA 2185  
Qy 364 GluGlnThrLysAspIleAsnAsnIleIlePheGlnIleGlnProMetAsnAsnThr 383  
Db 2186 TTAGAAATACCTGATGAGGAAATCTAAGTCAAGTACTAGTATGAAATCAAAATATGAT 2245  
Qy 384 TyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIle 403  
Db 2246 CAAGGAATAACAAATAAATGTC---AAATGAATTTACAAGATAATAATGGGAATGATATA 2302  
Qy 404 CysSerIleGlyThrTyrArgPheArgLeuGlyAspTrpTyrArgHisAsnTyrLeu 423  
Db 2303 GCCTTTATAGATTTTCATCAGTTT-----AATAATATA 2335  
Qy 424 ValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440  
Db 2336 GCTAAAGTAGTAGCAAGTAATTTGGTATAATAGACAAATAGAAAGATCTAGTAGGCTTTG 2395  
Qy 441 ---ThrHisTrpGlyPheValProValSerGlu 450

Db 2396 GGTGCTCATGGGAATTTTCTCTGTAGATGAT 2428  
RESULT 12  
US-10-051-952-3  
; Sequence 3, Application US/10051952  
; Patent No. US20020107199A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, Patricia  
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
; FILE REFERENCE: 2933CIP  
; CURRENT APPLICATION NUMBER: US/10/051,952  
; CURRENT FILING DATE: 2002-01-17  
; PRIOR FILING DATE: 2000-12-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 3876  
; TYPE: DNA  
; ORGANISM: botulinum toxin  
US-10-051-952-3  
Alignment Scores:  
Pred. No.: 9,99e-54 Length: 3876  
Score: 561.00 Matches: 159  
Percent Similarity: 50.21% Conservative: 84  
Best Local Similarity: 32.85% Mismatches: 173  
Query Match: 23.46% Indels: 68  
DB: 12 Gaps: 21  
US-09-910-186A-10 (1-450) x US-10-051-952-3 (1-3876)  
Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsnGlu 22  
Db 2506 ATGCGTTTCATCTTCAATATATACCAATGATACAACTACTACTAGAAATGTTAATAAA 2565  
Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGluAsnArgLysAsnThrLeu 42  
Db 2566 TATAATAGCGAAATTTTAAATAATATATCTTAAATTAAGATATAAGGATATAATTTA 2625  
Qy 43 ValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnPro 62  
Db 2626 ATAGATTATCAGGATATGGGCAAGGTAGAGGTATATGATGGAGTCGAGCTTAATGAT 2685  
Qy 63 IlePheProPheAspPheLysLeuGlySerSerGlyLysArgGlyLysValIleVal 82  
Db 2686 -----AAAAATCAATTTAAATTAAGTACTAGTTCAGCA-----AATAGTAGATAGAGTG 2733  
Qy 83 ThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrp 102  
Db 2734 ACTCAAAATCAGATATACATATTTAATAGTGTCTCTTCTGATTAGCGTTAGCTTTGG 2793  
Qy 103 IleArgIleAsnLysTrp-----ValSerAsnLeu-----ProGlyTyrThr 116  
Db 2794 ATAAGAATACCTTAATATAAGAATGATGGTATACAAATATATATTCATATGAATATACA 2853  
Qy 117 IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeu 136  
Db 2854 ATATTAATCTGTAAGAAATAATATTCGGCTGGAATAATCTATTAGGGGTAAATAGATA 2913  
Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
Db 2914 ATATGGACTTTAATTGATATAAATGGAATAACCAAAATCGGTATTTTGAATATAACATA 2973  
Qy 157 SerAsnAlaProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMet 175  
Db 2974 AGAAGATATATCAGAGTATATAAATAGATGGTTTGTCTACTACTACTATAAT--- 3030  
Qy 176 MetGlyAsnMetLysIleTyrIleAsnGlyLysLeuLeuAspThrIleLysValLysGlu 195  
Db 3031 TTGATAACGCTAAATTTATATTAATGTTAGGTAGTAGAATCAATCAATACAGATATAAAGAT 3090

```
QY 196 LeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThr 215
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3091 ATAAGAGAAGTATTCTGTAATGCTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 3135
QY 216 GlyLeuIleThrSerAspSerAsnIleAsn---MetTrpIleArgAspPheThrIle 234
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3136 -----CGTGATATAGATAGAACACAACTTATTGGATGAATATTTCAGTATT 3183
QY 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnThr 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3184 TTTAATACGGAATTAAGTCAATCAATATTAAGAAAGATATAAAATCAATCATATAGC 3243
QY 255 AsnValIleValLysThrIleGlyAsnAspLeuArgTyrAsnLysGlyTyrMetVal 274
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3244 GAATATTAAAGATTTTGGGAATCTTTAATGTAGATTAAGAAATATATATGTTT 3303
QY 275 AsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer-----ArgGlnIle 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3304 AATCGGGGAATAAAATTCATATATAAACTAAAGAAAGATTACCTGATAGTGAAT 3363
QY 291 ValPheAsnThrArgArgAsnAsnAsn-----AspPheAsnGlu 303
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3364 TTAACACGTAGCAATATATATCAAAATCTTAATATATAATATAAGATTTATATATT 3423
QY 304 GlyTyrLysIleIleLysArgIleArgGlyAsnThr-----AsnAspThrArg 320
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3424 GGAGAAAAATTTATTATAAGAAG---AAGTCAAAATCTCAATCTATAAATGATGATATA 3480
QY 321 ValArgGlyLysIleLeuTyrPheAsp---MetThrIleAsnAsnLys----- 336
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3481 GTTGAAGAAAGATATATATATCTAGATTTTAAATTTAAATCAAGAGTGGAGAGTA 3540
QY 337 ---AlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGlu 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3541 TATACCTATAAATTTTAAAGAAAGAGAA-----GAA 3573
QY 356 AspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePhe 375
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3574 AATGTGTTTACCTCTTAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 3624
QY 376 GlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGln---IlePheLysSerAsn 394
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3625 CAATAAAGAAATATGATGACAGCAACATATATTTGTGAGTGTGTTTAAAAA--- 3681
QY 395 PheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGly 414
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 3682 ---GATGAGAAAGTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3738
QY 415 Gly-----AspTyrTyrArgHisAsnTyrLeuVal-----ProThr 426
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3739 GGAATTGATTGAAGAGTATAAAGATTTATTTGTATAGTAATGGTAAAGAG 3798
QY 427 ValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHisTrpGlyPheVal 446
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3799 GTAAAGGAACCATATAATTTAAATTC-----GGATGTAATGGCGATTTAT 3849
QY 447 ProValSerGlu 450
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3850 CCTAAGATGAA 3861
RESULT 13
US-09-350-756-1
; Sequence 1, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
```

```
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 1
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-1

Alignment Scores:
Pred. No.: 5,35e-54 Length: 1338
Score: 557.50 Matches: 144
Percent Similarity: 48.80% Conservative: 80
Best Local Similarity: 31.37% Mismatches: 180
Query Match: 23.32% Indels: 55
Db: 10 Gaps: 11

US-09-910-186a-10 (1-450) x US-09-350-756-1 (1-1338)
QY 16 LeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerIleu 35
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 18 CTGCTGCTACCTTCACTCAATACATCAAGAACATCATCAATACCTCATCATCTGAACCTG 77
QY 36 GlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlu 55
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 78 CGTAGCATCCATCACTGATCGACCTGCTCGCTCGCTTCCAGTTCCTCAAAATCAACATCGT 137
QY 56 GlyAspValGlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySer 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 138 TCTAAAGTTAACTTCGATCCGATCGACAAAGATCAATCCAGCTGTTCAATCTGGAATCT 197
QY 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAsnSer 92
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 198 TCC-----AAATCGAAGTTATCTCGAAGATGCTATCGTATACACTCT 242
QY 93 MetTyrGluSerPheSerIleSerPheThrPheIleArgIleAsnLysTrpValSerAsnLeu 112
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 ATGTAGAAAACTTCTCCACCTCTCTGATCCGCTATCCGAAATACCTTCAACTCCATC 302
QY 113 Pro-----GlyTyrThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIle 129
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 303 TCTCTGAACAAATGAATACACCATCACTCACTCGTGAAGAACAACTTCGGTGGAAAGTA 362
QY 130 GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer 149
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 363 TCCTGAACATACGGTGAATCATCTGCACTCTGCAGACACTCAGGAAATCAACAGCGT 422
QY 150 IleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr---AsnLysTrpPhePhe 168
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 423 GTTGATTAATCAATCTCATGATGATCAACATCTCTGACATCATCATCAATCGCTGATC 482
QY 169 ValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeu 188
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 GTTACCATCACCAACATCTGCTGAATACTCCAAATCTACATCAACGCGCTCTGATC 542
QY 189 AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGlu 208
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 GACGAGAAACCGATCCCAATCTGGGTAACTCCAGCTTCTTAATAACATCATGTTCATA 602
QY 209 IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrp 228
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 603 CTGACCGTTGCTGTCGACT-----CACGCTACATCTGG 638
QY 229 IleArgAspPheThrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 639 ATCAATACTTCAATCTGTTTCGACAAAGAACTGACGAGAAAAAATAACAGACCTGTAC 698
```



```
QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362
Db 3580 -----GCTACTAATCATCACAGGCGGTAGAAAAAATACTAAGT 3621
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382
Db 3622 GCATTAGAAATACCTGATAGAAATCTAAGTCAAGTAGTAGTAATCAAGTCACAAAAAT 3681
QY 383 ThrTyrTyrThrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402
Db 3682 GATCAAGGAATAACAATAAATGC---AAATGAATTTACAGATAAATAATGGGAATGAT 3738
QY 403 IleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHisAsnTyr 422
Db 3739 ATAGGCTTTATAGGAATTCATCAGTTT-----AATAAT 3771
QY 423 LeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer----- 440
Db 3772 ATAGCTAAACTAGTACAGATTAATGGTATAATAGACAAATAGAAAGATCTAGTAGGACT 3831
QY 441 -----ThrHisTrpGlyPheValProValSerGlu 450
Db 3832 TTGGGTGCTCATGGGAATTTATCTCTGTAGATGAT 3867

RESULT 15
US-09-350-756-9
; Sequence 9, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 9
; LENGTH: 1327
; TYPE: DNA
; ORGANISM: Clostridium botulinum
; FEATURE:
US-09-350-756-9

Alignment Scores:
Pred. No.: 3,75e-53 Length: 1327
Score: 550.00 Matches: 146
Percent Similarity: 48.39% Conservative: 80
Best Local Similarity: 31.26% Mismatches: 174
Query Match: 23.00% Indels: 67
DB: Gaps: 12

US-09-910-186a-10 (1-450) x US-09-350-756-9 (1-1327)
QY 16 LeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeu 35
Db 10 TTATATCTACATTTACTGAAATATATAGAAATATTATTAATACCTCTCTATATGAATTGA 69
QY 36 GlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGlu 55
Db 70 AGATATGAAGTAACTATCTTATAGACTTATCTAGGTATGCATCAAAAATAAATATTGGT 129
QY 56 GlyAspValGlnLeuAsnProIlePheProPheAsp-----PheLysLeuGlySer 72
Db 130 AGTAAAGTAAATTTTGATCCAAATAGATAAAAAATCAAATTCATTTATTTAATTAGAAGT 189
```

```
QY 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyrAspSer 92
Db 190 AGT-----AAATTCAGGTAAATTTTAAAAATGCTATTTGTATATAATAGT 234
QY 93 MetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpValSerAsnLeu 112
Db 235 ATGTATGAAATTTTGTACTACTGCTTTTGGATAAGAAATCTCTAAGTATTTTAAACAGTATA 294
QY 113 Pro-----GlyTyrThrIleIleAspSerValLysAsnAsnSerGlyTrpSerIle 129
Db 295 AGTCTAAATTAATGAATATACATATAATTAATGTATGGAAATAATATTCAGGATGGAAGTA 354
QY 130 GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer 149
Db 355 TCACCTTAATTTATGGTGAATAATCTGGACTTTACAGGACTACTCAGGAAATAAACAAGA 414
QY 150 IleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyr---AsnLysTrpPhePhe 168
Db 415 GTAGTTTAAATATACAGTCAATGATTAATATATATCATGATTAATAACAGAGATGATTTT 474
QY 169 ValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIle 188
Db 475 GTAACATCATCAATAATAGATAAATCTAAATTTATATAAATGGAAGATTAATA 534
QY 189 AspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGlu 208
Db 535 GATCAAAACCAATTTCAATTTAGTAAATATTCAGTGTAGTAATAATATATATGTTTAA 594
QY 209 IleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrp 228
Db 595 TTAGATGGTGTAGAGATACA-----CATAGATATATATTGG 630
QY 229 IleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe 248
Db 631 ATAAATATTTTAAATCTTTTGTATAGGAATTAATGAAAAGAATCAAGAAATTTATAT 690
QY 249 AsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAsnAspLeuArgTyrAsn 268
Db 691 GATAATCAATCAAAATTCAGGTATTTTAAAGACTTTTGGGTGATTTATTTACAATGAT 750
QY 269 LysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSer--- 287
Db 751 AAACCATACTATATATGTTAAATTTATATGATCCAAATAATATGTCGATGTAATAATGTA 810
QY 288 -----ArgGlnIleValPheAsnThrArg 296
Db 811 GGTATTAGAGTTATATGTTCTTAAAGGCGCTAGAGTAGCGTAAATGACTACA----- 864
QY 297 AsnAsnAspPhe-Asn-----GluGlyTyrLysIleI 308
Db 865 -----AATGACTACAACATTTATTTAAATTCAGTTTGTATAGGGGACAAAATTTAT 918
QY 308 eileLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTy 328
Db 919 TATAAAAAATATGCTTCGGAATAAAGATAATATTTGTAGAAATAATGATCGTGATA 978
QY 328 rPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTy 348
Db 979 TATTAATGTAGTAGTTTAAATAAAGAAATATAGGTTA----- 1015
QY 348 rAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAs 368
Db 1016 -GCTACTATGCTACACAGCGCGGTAGAAAAAATACTAACTGCGATAGAAATACCTGA 1074
QY 368 pIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSe 388
Db 1075 TGTAGGAAATCTAAGTCAAGTAGTAGTAATGAAGTCAAAAAATGATCAAGGAATAACAAA 1134
QY 388 rGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyTh 408
Db 1135 TAAATGC---AAATGAATTTACAGATAATATATGGGAATGATAGGCTTTATAGGATT 1191
QY 408 rTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsnTyrLeuValProThrVally 428
```

```

Db 1192 TCATCAGTTT-----|||   |||   |||
QY 428 scInGlyAsnTyRAlaSerLeuLeuGluSerThrSer-----AATAATATAGCTAAACTAGTAGC 1224
Db 1225 AAGTAATTGGTATATAGACAAATAGAAAGATCTAGTAGGACTTTGGGTTGCTCATGGGA 1284
QY 444 yPheValProValSerGlu 450
Db 1285 ATTTATTCTGTAGATGAT 1303

```

Search completed: November 7, 2002, 20:06:56  
 Job time : 86 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 7, 2002, 17:48:22 ; Search time 3151 Seconds  
(without alignments)  
3590.644 Million cell updates/sec

Title: US-09-910-186A-10

Perfect score: 2391

Sequence: 1 MTTFNFTSTNNLLKDII.....NYASLLESTHWGVPVNSE 450

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09910186/runat\_04112002.111618.20892/app\_query.fasta.1.647  
-DB=Pending\_Patents\_NA\_Main -QFMT=fastap -SUFFIX=rnrm -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MNLLEN=0 -MAXLEN=2000000000  
-USER=US09910186.ecgn.1.1.1396 -runat\_04112002.111618.20892 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Pending\_Patents\_NA\_Main :

1: /cgn2\_6/ptodata/1/pna/PCTUS\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*

29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq.\*  
44: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6032\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6033\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6034\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6035\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6036\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6037\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6038\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6039\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6040\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6041\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2391	100.0	1371	23	US-09-611-419A-9
2	2391	100.0	1371	34	US-09-910-186A-9
3	2386	99.8	3876	1	PCT-US97-15394-59
4	2386	99.8	3876	11	US-08-704-159-59
5	2386	99.8	3876	42	US-10-205-516-5
6	2386	99.8	3876	43	US-10-271-012-59



```
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
Db 790 TGGGGTAAGGACCTGAGATACAAAGAGTACTACATGGTCAACATCGACTACTTGAAC 849
QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300
Db 850 AGATACATGTAGCCCAAGTCCAGACAGATCTCTCTACACACAGACGTAACACACGAC 909
QY 301 PheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320
Db 910 TTCACGAGGGTTACAGATCATCATCAAGGTATCAGAGGTACACACACGACACGAGA 969
QY 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
Db 970 GTCAGAGGTGGTACATCTGTCTACATGACATGACTATCAACACAAAGGCGCTACAACCTG 1029
QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1030 TTCATGAGAAGACGACCATGATGACCCGACCAACCATCCACCGAGGACATCTAGCGCATC 1089
QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380
Db 1090 GGTCTGCGTGCAGACACCAAGACATCAACACACATCATCTTCCAGATCCACGCAATG 1149
QY 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1150 AACACACTTACTACTACGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATC 1209
QY 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHis 420
Db 1210 TCGGTPATCTGTTCATCGGTACCTACAGATTCGGTCTGGTGGTGACTGTGACAGACAC 1269
QY 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
Db 1270 AACTACTTGGTTCCAACTCTCAAGCAGGTAACACGCTCTCTTGGAGTCCACATCC 1329
QY 441 ThrHisTrpGlyPheValProValSerGlu 450
Db 1330 ACCCACTGGGATTCGTCACAGTCTCCGAG 1359

RESULT 2
US-09-910-186A-9
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; NAME/KEY: CDS
; LOCATION: (10)...(1359)
US-09-910-186A-9

Alignment Scores:
Pred. No.: 4,07e-230 Length: 1371
Score: 2391.00 Matches: 450
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 34 Gaps: 0

US-09-910-186A-10 (1-450) x US-09-910-186A-9 (1-1371)
QY 1 MetThrIlePhePheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle 20
Db 10 ATGACATCCCATTCACATCTTCTCTACACCAACAACTCCCTGTTGAAGGACATCATC 69
QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
Db 70 AACGAGTACTTCAACACATCAACGACTCCAAAGATCCTGTCCTGCAGAACCGTAAGAAC 129
QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
Db 130 ACCTTGGTGCACACTCCGGTTACAACCGCAGGTCTCCGAGGAGGTGACGCCAGCTG 189
QY 61 AsnProIlePhePheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80
Db 190 AACCCAACTTCTCCATTCGACTTCAAGCTGGTTCCTCCGCTGAGGACAGAGTAAAGTC 249
QY 81 IleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
Db 250 ATCGCACCCAGACAGACATCGTCTACAACTCCATGTACGAGTCTCTTCCATCTCC 309
QY 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleLeuAspSer 120
Db 310 TTCCTGGATCAGATCAACAGTGGTCTCCAACTTGCAGGTACACCATCATCGACTCC 369
QY 121 ValLysAsnAsnSerGlyTrpSerIleGlyIleLeuSerAsnPheLeuValPheThrLeu 140
Db 370 GTCAAGACACACTCCGGTGGTTCATCGGTATCATCTCCAACTTCTTGGTCTCCACCTG 429
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAla 160
Db 430 AAGCAGACGAGGACTCCGAGCGATCCATCAACTTCTCTACGACATCTCCAAACAGCT 489
QY 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
Db 490 CCTGGTTACAAACAGTGGTCTTCGTCACCGTCAACCAACATGATGGTAAATGAAG 549
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyLeuAsn 200
Db 550 ATCTACATCAACGGTAAGTGTATCGACACCATCAAGTCAAGGATGACCGGTATCAAC 609
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 610 TTCTCCAGACCATCACCTTCGAGATCAACAAAGATCCCAAGTCCACCGGTCTGATCACCTCC 669
QY 221 AspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240
Db 670 GACTCGACAAACATCAACATGTGGATCGGTGACTTCTACATCTTCGCCAAGGATGGAC 729
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyr 260
Db 730 GGTAAAGACATCAACATCCTGTTCAACTCCTTGCAGTACACCAACGCTCGTCAAGGACTAC 789
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
```

## RESULT 2

```
US-09-910-186A-9
; SEQUENCE 9, Application US/09910186A
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research & Materiel Command
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; FILE REFERENCE: A33626-A 067252, 0107
; CURRENT APPLICATION NUMBER: US/09/910,186A
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-05
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,873
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1371
```

```
Db 790 TGGGTAACGACCTGAGATACAAACAGGAGTACTACATGCTCAACATCGACTACTGTGAC 849
Qy 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsp 300
Db 850 AGATACATGTACGCCACTCCAGACAGATCGTCTTCAACACGAGAGTACACACACGAC 909
Qy 301 PheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArg 320
Db 910 TTCACAGGGGTTACAAAGATCATCATCAAGCGTATCAGAGGTAAACACACACACACCA 969
Qy 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
Db 970 GTCAGAGGTGTGACATCCCTGTACTTCGACATGACTATCAACACAGGCCCTACACCTG 1029
Qy 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
Db 1030 TTCATGAAGACGAGACCATGTACGCCGACACCACTCCACCGAGGACATCTACGCCATC 1089
Qy 361 GlyLeuArgGluGlnThrLysAspIleAspAsnIleIlePheGlnIleGlnProMet 380
Db 1090 GGTCTGGTGACGACAGACAGGACATCAACGACACATCATCTCCAGATCCAGCCAAAG 1149
Qy 381 AsnAsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
Db 1150 AACACACCTACTACTACGCTTCCAGATCTTCAAGTCCAACTCAACGGTGAGAACATC 1209
Qy 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyClyAspThrArgHis 420
Db 1210 TCCGGTACTCTTCCATCCGCTACTACAGATCCGCTGGTGAGTGGTGGTGGTGGTGGT 1269
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
Db 1270 AACTACTGTGTCCAACTGCAACAGCGGTAACTACGCCCTCTGTGGTGCCACTTCC 1329
Qy 441 ThrHisTrpGlyPheValProValSerGlu 450
Db 1330 ACCACTGGGATTCGCTCCAGTCTCCGAG 1359

RESULT 3
PCT-US97-15394-59
; Sequence 59, Application PC/TUS9715394
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/15394
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3876 base pairs.
```

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3873
PCT-US97-15394-59

Alignment Scores:
Pred. No.: 4,77e-229 Length: 3876
Score: 2386.00 Matches: 449
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 1 Gaps: 0

US-09-910-186a-10 (1-450) x PCT-US97-15394-59 (1-3876)
Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleAsn 21
Db 2327 ACAATACCCCTTAAATATTTTTCATATACTAATAATCTTTTAAAGATATAATTAAT 2585
Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2587 GAATATTTCAATATATATATGATTCAAAAATTTTTCAGCCTCAAAAACAGAAAAATACT 2646
Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 2647 TTAGTGATACATCAGGATATAATGCAAGTGTAGTGAAGAGGCGATGTTCCAGCTTAT 2706
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2707 CCAATATTTCCATTTGACTTTAAATTAGTGTAGTTTCAGGGGAGGATAGAGTAAAGTTATA 2766
Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101
Db 2767 GTAACCCAGAAATATGTATATAATTTCTATATGTAAGAGTTTATGCAATTAGTTT 2826
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
Db 2827 TGGATTAGATAATAATAATGGTAACTAATTTTACCCTGGATATATAATTCATAGTGT 2886
Qy 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
Db 2887 AAAATAAATCAGGTTGGAGTATAGTATATATAGTAATTTTATTTAGTATTTACTTAAA 2946
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaPro 161
Db 2947 CAATAATCAAGATAGTAGAACAAAGTATAAATTTTAGTATATATATCAATAATGCTCT 3006
Qy 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3007 GGATACAAATAATGGTTTTTTTGTAACTGTTTACTTAACAATATGATGGAAATATGAAGAT 3066
Qy 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3067 TATATAATGAAATTAATTAATAGATATATAAAGTTAAAGAACTAACTGGAATTAATTT 3126
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3127 AGCAAAACTATACATTTGAATAATAATAAATTTCCAGATACCGGTTTGATTTACTTCAT 3186
Qy 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241
Db 3187 TCTGATACATCAATATGTGGTAAGAGATTTTATATATTTTCTTAAGAAATAGATGT 3246
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261
Db 3247 AAGATATATATATATTAATTAATAGCTTCAATATACATTAATGTTGTAAGAAATATTGG 3306
Qy 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281
```

```
Db 3307 GGAATGATTTAAGATATATAAAGAAATATATATGTTAAATATATAGTATTTAAATAGA 3366
Qy 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301
Db 3367 TATATGATGGAACCTCAGCAGAAATGTTTTTAAATACACGTAGAAATATATGACTTC 3426
Qy 302 AsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
Db 3427 AATGAAGATATAAAATATATAAAGAAATCAGAGAAATACAAATGATAGAGTA 3486
Qy 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341
Db 3487 CGAGGAGAGATATTTATATTTGATGATGACATTAATTAACAACGATATATTTGTTT 3546
Qy 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361
Db 3547 ATGAAGATGAACATATGTATGCAGATTAATCAATCAAAATTTTAAATGAGAAATATTTCT 3606
Qy 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
Db 3607 TTAGAGAACAAACAGGATATAAATGATATATATATATTTCAATACACCAATGAAT 3666
Qy 382 AsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
Db 3667 AATACTTATATACGATCTCAATATATTAATCAAAATTTTAAATGAGAAATATTTCT 3726
Qy 402 GlyIleCysSerIleGlyThrTyrArgPheArgIleGlyAspThrTyrArgHisAsn 421
Db 3727 GGAATATGTTCAATAGGACTATTCGTTTATAGCTGGAGGTGATTTGATATACACAT 3786
Qy 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
Db 3787 TATTTGGTGCCTACTGTGAAGCAAGAAATATGCTTCATATATAGATCAACATCAACT 3846
Qy 442 HisTrpGlyPheValProValSerGlu 450
Db 3847 CATTTGGGTTTGTACCTGTAAAGTAA 3873

RESULT 4
US-08-704-159-59
; Sequence 59, Application US/08704159
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08704,159
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPFD-02304
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3876 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3873
US-08-704-159-59

Alignment Scores:
Pred. No.: 4,77e-229 Length: 3876
Score: 2386.00 Matches: 449
Best Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.79% Indels: 0
DB: 11 Gaps: 0

US-09-910-186A-10 (1-450) x US-08-704-159-59 (1-3876)
Qy 2 ThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21
Db 2527 ACAATACCCCTTAAATATTTTTCATATACATAATAATCTTTATTAAGATATAATAT 2586
Qy 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
Db 2587 GAATATTTCAATATATATATGATTCAAAAATTTTGAGCCTACAAACACAGAAAAATAC 2646
Qy 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
Db 2647 TTATGATACATCAGGATATATATGCAGAAAGTGCAGAAAGCGGATGTTGACCTTAAT 2706
Qy 62 ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
Db 2707 CCAATATTTCCATTTGACTTTAAATTAGGTAGTTTCAGGGGAGGATAGAGTAAAGTTATA 2766
Qy 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrCysSerPheSerIleSerPhe 101
Db 2767 GTAACCCAGAAATGAAATATTTGATATATATTTTATGATGAAAGTTTTCAGCATAGTTT 2826
Qy 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121
Db 2827 TGGATTAGATAATAATAAATGGTAAAGTAATTTTACCCTGGATATATCTAATATGATAGTT 2886
Qy 122 LysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnAspPheLeuValPheThrLeuLys 141
Db 2887 AAAATAACACAGGTGGAGTATAGGTATTTAGTAATTTTATGATTTTCTTTTAAAT 2946
Qy 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161
Db 2947 CAAATGAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAAAATAATGCTCT 3006
Qy 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 3007 GGATACAAATAAATGTTTTTTTGTAACTGTTTACTAAACAATATGAGGAATATGAAAGAT 3066
Qy 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
Db 3067 TATATAATGGAATTAATAGATACATATAAAGTTAAAGAACTAACTGGAATTAATTTT 3126
Qy 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
Db 3127 AGCAAACTATAACATTTGAATAAATAAATTTCCAGATACCGGTTTGATTTCTCAGAT 3186
Qy 222 SerAspAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241
Db 3187 TCTGATACATCAATATGCGGATAGAGATTTTATATATTTCTTAAGAAATAGATGCT 3246
Qy 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrp 261
Db 3247 AAAGATATATATATATTTAATAGCTTCAATATCTAATAATGTTTGTAAAGATTTATCG 3306
Qy 262 GlyAsnAspLeuArgTyrAsnLysGluTyrMetValAsnIleAspTyrLeuAsnArg 281
```

Score:	2386.00	Matches:	449
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.79%	Indels:	0
DB:	42	Gaps:	0
US-09-910-186A-10 (1-450) x US-10-205-516-5 (1-3876)			
QY	2	ThrIleProPheAsnIlePheSerTyThrAsnAsnSerLeuLeuLeuLysAspIleIleAsn	21
DB	2527	ACAATACCCCTTAATATTTTTCATATACTATAATCTTTTAAAGATATATATTAAT	2586
QY	22	GLuTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr	41
DB	2587	GAATATTTCAATAATATTAATGATTCAAAAATTTTGACCTACAAAAACAGAAAAAATACT	2646
QY	42	LeuValAspThrSerGlyTyrrAsnLaGluValSerGluGluGlyAspValGlnLeuAsn	61
DB	2647	TTAGTGGATACATCAGATATATGACGAAGTGTAGTGAAGAGGCCATGTTACGCTTAAT	2706
QY	62	ProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle	81
DB	2707	CCAATATTTCCATTTCCACTTTAAATAGGTAGTTCTCAGGGGAGGATAGAGTAAAGTTATA	2766
QY	82	ValThrGlnAsnGluAsnIleValTyrrAsnSerMetTyrrGluSerPheSerIleSerPhe	101
DB	2767	GTAACCCAGAAATGAAATATGTATATATATTCATGTATGAAAGTTTTCAGCATTTAGTTT	2826
QY	102	TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleIleAspSerVal	121
DB	2827	TGATTTAGATAATAATAATGGGTAAAGTATTTTACCTGGATATACTATATTAATGATAGTGT	2886
QY	122	LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys	141
DB	2887	AAAAATPACTCAGTTGGAGTAGTAGTATTAATAGTAATTTTAGTATTTTACTTTAAAA	2946
QY	142	GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrrAspIleSerAsnAsnAlapro	161
DB	2947	CAAAATGAAGATAGTCAACAAAGTATAAATTTTAGTTATGATATATCAAAATAATGCTCCT	3006
QY	162	GlyTyrrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle	181
DB	3007	GGATACAAATAAATGGTTTTTGTACTGTTACTACAAATATGATGGGAAATATGAAGATT	3066
QY	182	TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe	201
DB	3067	TATATAATGGAAAAATATAGATACATATAAAGTATAAAGAACTAACCGGAATTAATTTT	3126
QY	202	SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp	221
DB	3127	AGCAAAACTATACATTTGAAATAAATAAATCCAGATACCGGTTGTTACTTCTCAGAT	3186
QY	222	SerAspAsnIleAsnMetTrpIleArgAspPheTyrrIlePheAlaLysGluLeuAspGly	241
DB	3187	TCGTATCAACTCAATATGTGGATAAGAGATTTTATATATTGCTAAAGAATTAGATGGT	3246
QY	242	LysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValValLysAspTyrrTrp	261
DB	3247	AAAGATATTAATATATTTAATACCTTGCATATACTAATGTTGTTAAAAGATTATTGG	3306
QY	262	GlyAsnAspLeuArgTyrrAsnLysGluTyrrTyrrMetValAsnIleAspTyrrLeuAsnArg	281
DB	3307	GGAAATGATTTAAGATATATAAAGAATAATATATGTTTAAATAGATTTATTAATAAGA	3366
QY	282	TyrMetTyrrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPhe	301
DB	3367	TATATGATCGGAACTCAGACAAAATGTTTTTAATACACGTAGAAAAATAAATGACATTC	3426
QY	302	AsnGluGlyTyrrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal	321
DB	3427	AATGAAGGATATAAATTTAATAAAGAAAGATCAGAGGAATAATCAAAATGATACTAGAGTA	3486
QY	322	ArgGlyGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsnLeuPhe	341

Qy 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnIysAlaTyrAsnLeuPhe 341



Db	3427	AAATGAGGATATAAAATTAATAAAAAAGATCAGAGGAAATACAAATGGATACTAGAGTA	3486
Qy	322	ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe	341
Db	3487	CGAGGAGGAGATATTTTATATTTTGATGACAAATTAATAACAAGACATATAATTGCTTT	3546
Qy	342	MetIleAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly	361
Db	3547	ATGAAGAATGAAACTATGTATGCAGATAATCATCTACTGAAGATATATATGCTATAGGT	3606
Qy	362	LeuArgGluGlnThrLysAspIleAsnAspIleIlePheGlnIleGlnPrometAsn	381
Db	3607	TTAAGAGAACAAACAAGGATATAATGATATATATATTTCATTAATACAAACCAATGAAT	3666
Qy	382	AsnThrTyrTyrThrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer	401
Db	3667	AAVACTATATTATTCGCATCTCAAAATATTAAATCAAAATTTAATGGAGAAATATTCT	3726
Qy	402	GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTptTyrArgHisAsn	421
Db	3727	GGATATGTTCAATAGTACTTATCGTTTTTAGACHTGGAGGTGATGGGTATAGACAAAT	3786
Qy	422	TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr	441
Db	3787	TATTTGGTGCTACTCTGAAGCAGGAATTTATGCTTCATTATAGAATCAACATCACT	3846
Qy	442	HisTrpGlyPheValProValSerGlu	450
Db	3847	CATTGGGGTTTTTGTACCTGTAAAGTAA	3873

## RESULT 7

```

US-10-205-516-19
: Sequence 19, Application US/10205516
: GENERAL INFORMATION:
: APPLICANT: Zhong, Jun
: TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum
: TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique
: FILE REFERENCE: Jzbtxl
: CURRENT APPLICATION NUMBER: US/10/205,516
: CURRENT FILING DATE: 2002-07-25
: NUMBER OF SEQ ID NOS: 28
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 19
: LENGTH: 3906
: TYPE: DNA
: ORGANISM: Clostridium botulinum
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(3906)
: FEATURE:
: NAME/KEY: misc_signal
: LOCATION: (1347)..(1358)
: OTHER INFORMATION: factor Xa site
: FEATURE:
: NAME/KEY: misc_signal
: LOCATION: (3886)..(3903)
: OTHER INFORMATION: 6-histidine tag
US-10-205-516-19

```

Alignment Scores:	4.82e-229	Length:	3306
Pred. No.:	2386.00	Matches:	449
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	99.79%	Gaps:	0
DB:	42		

US-09-910-186A-10 (1-450) x US-10-205-516-19 (1-3906)

Qy	2	ThrIleProPheAsnIlePheSerTyrThrAsnAspSerLeuLeuLysAspIleAsn	21
Db	2539	ACRAATACCCCTTTATATATTTTTTCATATACATAATATCTTTATTAAAGATATAATATAT	2598



QY 382 AsnThrTyrTyrAlaSerGlnIlePheIysSerAsnPheAsnGlyGluAsnIleSer 401  
Db 3679 AATACATATATATAGCATCTCAAAATATATTAATCAAAATTTTAATGGAGAAATATTTCT 3738  
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHisAsn 421  
Db 3739 GGAATATGTTCAATAGGTACTATCGTTTAGACTTGGAGGTGATTGGTATAGACACAAT 3798  
QY 422 TyrLeuValProThrValIysGlnGlyAsnTyrAlaSerLeuLeuGluSerThr 441  
Db 3799 TATTTGGTCCCTACTGTGAAGCAAGAAATATGCTTCATTTAGTAATCAACATCACT 3858  
QY 442 HisTyrGlyPheValProValSerGlu 450  
Db 3859 CATTTGGGTTTGTACCTGTAACTGAA 3885  
RESULT 8  
US-08-954-302-2  
; Sequence 2, Application US/08954302  
; GENERAL INFORMATION:  
; APPLICANT: Lance Simpson, Nikita Kiyatkin,  
; APPLICANT: Andrew Maksymowich  
; TITLE OF INVENTION: Compositions and Methods for Systemic  
; TITLE OF INVENTION: Delivery of Oral Vaccines and Therapeutic Agents  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Jane Massey Licata, Esq.  
; STREET: 66 E. Main Street  
; CITY: Marlton  
; STATE: NJ USA  
; COUNTRY: USA  
; ZIP: 08053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
; COMPUTER: IBM 486  
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/954,302  
; FILING DATE: herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane Massey Licata  
; REGISTRATION NUMBER: 32,257  
; REFERENCE/DOCKET NUMBER: JEFF-0164  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 779-2400  
; TELEFAX: (609) 810-1454  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3950  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: SINGLE  
; TOPOLOGY: LINEAR  
; ANTI-SENSE: NO  
US-08-954-302-2  
Alignment Scores:  
Pred. No.: 4,89e-229 Length: 3950  
Score: 2386.00 Matches: 449  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 99.79% Indels: 0  
DB: 13 Gaps: 0  
US-09-910-186a-10 (1-450) x US-08-954-302-2 (1-3950)  
QY 2 ThrIleProPheAsnIlePheSerTyrThrAsnSerLeuLeuLysAspIleAsn 21  
Db 2558 ACAATACCCCTTTAATATTTTTCATATACTATATAATCTTTTATTAAAGATTAATTAAT 2617

QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
Db 2618 GAATATTTCAATAATATTAATGATTTCAAAATTTTGAGCCCTACAAAAACAGAAAAATACT 2677  
QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
Db 2678 TTAGTGGATACATCAGGATATATGACAGAGTGAAGAGCGGATGTTCAGCTTAAT 2737  
QY 62 ProIlePheProPheAspPheLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
Db 2738 CCAATATTTCCATTTGACTTTAAATTTAGTTAGTTTCCAGGGGAGGATAGAGTAAAGTTATA 2797  
QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPhe 101  
Db 2798 GTAACCCAGATGAATAATTTGATATAATCTATGATGAAGAGTTTAGCATTAAGTTT 2857  
QY 102 TyrIleArgIleAsnLysTyrPheValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
Db 2858 TGGATTAGAATAAATAAATGGTAAAGTAATTTACCTGGATATACATAATGATAGTGT 2917  
QY 122 LysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
Db 2918 AAAATAAATCAGTTGGATAGTATATTTAGTAAATTTTGTAGTATTTTACCTTTAAA 2977  
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161  
Db 2978 CAAATGAAGATAGTGAACAAAGTATAATTTTAGTTATGATATATCAATAATGCTCCT 3037  
QY 162 GlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
Db 3038 GGATACAAATAAATGGTTTGTAACTGTACTAACAAATATGATGGGAAATATGAAGATT 3097  
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
Db 3098 TATATAATGGAAATTAATAGATATATAAGTTAAAGAACTAACTGGAATTAATTT 3157  
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProaspThrGlyLeuIleThrSerAsp 221  
Db 3158 AGCAAAACTATAACATTTGAAATTAATAAATTCAGATACCCGTTTCATTTACATCTCAGAT 3217  
QY 222 SerAspAsnIleAsnMetTyrPheArgPheTyrIlePheAlaLysGluLeuAspGly 241  
Db 3218 TCTGATACATCAATATGTTGGATAGAGATTTTATATATTTGCTAAGAAATTAGATGGT 3277  
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrp 261  
Db 3278 AAAGATATTAATATATTTAAATAGCTTGAATATATACTAATGTTGTAAGAGATTAATGG 3337  
QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
Db 3338 GGAATGATTTAAGATATATAAAGAAATTTATATGTTAATATAGATTTATTAATAGA 3397  
QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
Db 3398 TATATGATGCGAACTCAGCAAAATTTGTTTAAATACACGTAGAAATAATAATGACTTC 3457  
QY 302 AsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
Db 3458 AATGAAGGATATAAATAATATAAAGAAATCAGAGGAAATCAATGATACATAGAGTA 3517  
QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
Db 3518 CGAGGAGGAGATATTTTATTTTATGATGACAAATTAATAACAAGCATATAATTTGTTT 3577  
QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361  
Db 3578 ATCAAGATGAACATATGATGACATATATCATAGTACTAGGATATATATGCTTATAGGT 3637  
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381  
Db 3638 TTAAGAGACACAAAGGATATAATGATTAATATATATTTCAAAATACACCAATGAAT 3697

```
QY 382 AsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
DB 3698 AATACTTATTATACGATCTCAATATTTAAATCAAAATTTTAAATGAGAAATATTCT 3757
QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsn 421
DB 3758 GGAATATGTTCAATAGGTACTATCGTTTAGACTTGGAGGTGATTGGTATAGACAAAT 3817
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441
DB 3818 TATTGGTGCTACTGTGAAGCAAGGAATATGCTTCATTATTAATCAACATCACT 3877
QY 442 HisTrpGlyPheValProValSerGlu 450
DB 3878 CATTGGGTTTGTACCTGTAAGTAA 3904

RESULT 9
US-09-350-756-3
; Sequence 3, Application US/09350756
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 3
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-3

Alignment Scores:
Pred. No.: 4.13e-229 Length: 1371
Score: 2381.00 Matches: 448
Percent Similarity: 99.78% Conservative: 1
Best Local Similarity: 99.56% Mismatches: 1
Query Match: 99.56% Indels: 0
DB: 17 Gaps: 0

US-09-910-186a-10 (1-450) x US-09-350-756-3 (1-1371)
QY 1 MetThrIleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIle 20
DB 10 ATCAACATCCCATTCACATCTTCCTACACCAACACTCCCTCGTTGAAGGACATCATC 69
QY 21 AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsn 40
DB 70 AAGGAGTACTTCAACAACATCAACACTCAAGATCCTGCTCCCTGCAGAACCGGTAAAGAC 129
QY 41 ThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeu 60
DB 130 ACCTTGGTCGACACTCCGGTTTACAAACCCGAGGTCTCCGAGGAGGTGAGCTCCAGCTG 169
QY 61 AsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysVal 80
DB 190 AACCCAACTTCCCATTCGACTTCAAGCTGGTTCCTCCGCTGAGGACAGAGGTAAAGTTC 249
QY 81 IleValThrGluAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100
DB 250 ATCGTCACCCAGAGAGAACATCTGTACAACTCCATGTACGAGGTCTTCTCCATCTCC 309
```

```
QY 101 PheTrpIleArgIleAsnLysTyrValSerAsnLeuProGlyTyrThrIleIleAspSer 120
DB 310 TTCGGATCAGATCAACAAGTGGTCTCCAACTTCCAGGTACACCATCATCGACTCC 369
QY 121 ValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeuValPheThrLeu 140
DB 370 GTCAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCACCTTCCGTGCTTCCACCTG 429
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAla 160
DB 430 AAGCAGAACAGAGNCTCCGAGAGTCCATCAACTTCTCTACGACATCTCCAAACAAGCT 489
QY 161 ProGlyTyrAsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMetLys 180
DB 490 CCTGTTACAAACAAGTGGTCTTCGTCACGTCACCAACAACATGATGGGTAAACATGAAG 549
QY 181 IleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn 200
DB 550 ATCTACATCAACGGTAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAAC 609
QY 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyIleIleThrSer 220
DB 610 TTCCTCAAGACCATCACTTCGAGATCAACAAGATCCACACACCGGTCTGATCACTCC 669
QY 221 AspSerAspAsnIleAsnMetTyrIleArgAspPheTyrIlePheAlaLysGluLeuAsp 240
DB 670 GACTCCGACAAATATCAACATGTGGATCCGTTCATCTTACATCTTCCCAAGGAGTTGGAC 729
QY 241 GlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyr 260
DB 730 GGTAAGGACATCAACATCTTCAACTCTTGCAGTACACCAACCTCGTCAAGGACTAC 789
QY 261 TrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsn 280
DB 790 TGGGTAAACGACCTGAGATCAACAAGAGGTACTATGTTGATCAACATCTGACTACTTGAAC 849
QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAsp 300
DB 850 AGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACACGAGGTAAACAACAAGAC 909
QY 301 PheAsnGluGlyTyrLysIleIleIleIleIleIleIleIleIleIleIleIleIleIle 320
DB 910 TTCACAGAGGTTTCAAGATCATCATCAAGGTATCAGAGGTAAACCAACGACACCAAGA 969
QY 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
DB 970 GTCAGAGGTGGTGACATCTCTACTTCGACATGACTATCAACAACAAGGCTTACAACCTG 1029
QY 341 PheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIle 360
DB 1030 TTCATGAAGAAGACGACCATGTACCGGACAAACCACTCCACGAGGACATCTACGCCATC 1089
QY 361 GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMet 380
DB 1090 GGTCTGCGTGAGCAGACCAAGGACATCAACGACAAACATCATCTTCCAGATCCAGCCAATG 1149
QY 381 AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIle 400
DB 1150 AACAAACACTTACTACTACGTTCCTCCAGATCTTCAAGTCCCACTTCAACGGTGAGAACATC 1209
QY 401 SerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTyrTyrArgHis 420
DB 1210 TCCGATATCTGTTCATCGGTACCTACAGATTCGCTGCTGGTGGTACTGCTGACAGACAC 1269
QY 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
DB 1270 AACTACTTGGTTCCAACTGTCAAGCAGGTAACTACGCCCTCTTGTGGAGTCCACTTCC 1329
QY 441 ThrHisTrpGlyPheValProValSerGlu 450
DB 1330 ACCCACTGGGATTCGTCCTCCAGTCTCCGAG 1359

RESULT 10
```

```
US-09-730-237-4
; Sequence 4, Application US/09730237
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Treating Hyperhidrosis
; FILE REFERENCE: 2933
; CURRENT APPLICATION NUMBER: US/09730,237
; CURRENT FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3876
; TYPE: DNA
; ORGANISM: botulinum toxin
US-09-730-237-4
Alignment Scores:
Pred. No.: 7,83e-227 Length: 3876
Score: 2364.00 Matches: 447
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 2
Query Match: 98.87% Indels: 0
DB: 29 Gaps: 0
US-09-910-186a-10 (1-450) x US-09-730-237-4 (1-3876)
QY 2 ThrillePropheAsnIlePheSerTyrrAsnSerLeuLeuLysAspIleIleAsn 21
DB 2527 ACAATACCCCTTTAAATATTTTCATATACATAATATCTTTATTAAGAGATATAAT 2586
QY 22 GluTyrrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41
DB 2587 GAATATTTCCAAATATTAATGATTCAAAATTTTGAGCCCTACAAAACAGAAAATACT 2646
QY 42 LeuValAspThrSerGlyTyrrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61
DB 2647 TTATGGGATACATCAGGATATAATGACAGAGCTGAGTGAGAGAGCGATGTCAGCTTAAT 2706
QY 62 ProIlePhePropheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81
DB 2707 CCAATATTTCCATTTGACTTTAAATAGTAGTACGAGGAGGAGGAGGAGGAGGAGGAGG 2766
QY 82 ValThrGlnAsnGluAsnIleValTyrrAsnSerMetTyrrGluSerPheSerIleSerPhe 101
DB 2767 GTAACCCAGATGAATATGATATATATTTCTATGATGAAGATTTTAGCATTTAGTTT 2826
QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrrThrIleIleAspSerVal 121
DB 2827 TGGATTAGAATAAATAATGGTAAAGTAAATTTACCTGGATATACATAATGATAGTGT 2886
QY 122 LysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141
DB 2887 AAAATAACTCAGGTGGAGTATAGGTATTTATTTAGTATTTTATTTAGTATTTTAA 2946
QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrrAspIleSerAsnAlaPro 161
DB 2947 CAAATGAAGATAGTGAACAAGTATAATTTAGTTATGATATATCAATCAATATGCTCT 3006
QY 162 GlyTyrrAsnLysTrpPheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
DB 3007 GGATAGAATAAATGTTTTTTTGAATCTTACTACCAATATGATGGGAAATATGAGATT 3066
QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201
DB 3067 TATATAAATGGAATTAATAGATATATAAAGTTAAAGAACTAACTGGAATTAATTT 3126
QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221
DB 3127 AGCAAACTATACATTTGAAATAAATAAATTTCCAGATACCGGTTTGAATTCAGAT 3186
QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrrIlePheAlaLysGluLeuAspGly 241
DB 3187 TCTGATACATCAATATGCGGATAGAGATTTTATATATTTTGTATAGAAATATAGATGCT 3246
QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrrThrAsnValLysAspTyrrTrp 261
DB 3247 AAAGATATTAATATTAATTAATAGCTTGAATATATACTAATATTTGTAAGATATTGG 3306
QY 262 GlyAsnAspLeuArgTyrrAsnLysGluTyrrMetValAsnIleAspTyrrLeuAsnArg 281
DB 3307 GGAATGATTTAAGATATATAAAGATATATATGTTAATATAGATTTTAAATAGA 3366
QY 282 TyrMetTyrrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAspPhe 301
DB 3367 TATATGATGCAACTCAGCAAAATTTTAAATACAGTAAATATAATATGACTTC 3426
QY 302 AsnGluGlyTyrrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321
DB 3427 AATGAGGATATAAATATATAAATAAAGAAATCAGAGAAATACAAATGATAGATGA 3486
QY 322 ArgGlyGlyAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsnLeuPhe 341
DB 3487 CGAGGAGGAGATATTTATATTTTGATATGACAAATTAATAACAAAGCATATAATTTGTT 3546
QY 342 MetLysAsnGluThrMetTyrrAlaAspAsnHisSerThrGluAspIleTyrrAlaIleGly 361
DB 3547 ATGAAGAATGAAGACTATGATGACAGATATCAATAGTACTGAGATATATATGCTATAGT 3606
QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsn 381
DB 3607 TTAAGAGAACAAACAAAGGATATAATGATAATATATATTTTCAAAATACAAATGAAT 3666
QY 382 AsnThrTyrrTyrrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401
DB 3667 AATACCTTATTTAGGATCTCAATATTTAAATCAATTTTATGGAGAAATATTTCT 3726
QY 402 GlyIleCysSerIleGlyThrTyrrArgPheArgLeuGlyGlyAspTyrrArgHisAsn 421
DB 3727 GGAATATCTCAATAGGTACTTATCTGTTTAGACTTGGAGTGATTTGGTATAGACAAAT 3786
QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrrAlaSerLeuLeuGluSerThrSerThr 441
DB 3787 TATTTGGTCCCTACTGTGAACAGGAAATATGCTTATTTAGTAATCAATCAATCACT 3846
QY 442 HistTrpGlyPheValProValSerGlu 450
DB 3847 CATTTGGGCTTTTGACCTGTAAAGTGA 3873
RESULT 11
US-10-051-952-4
; Sequence 4, Application US/10051952
; GENERAL INFORMATION:
; APPLICANT: Walker, Patricia
; TITLE OF INVENTION: Methods of Administering Botulinum Toxin
; FILE REFERENCE: 2933CIP
; CURRENT APPLICATION NUMBER: US/10/051,952
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: 09/730,237
; PRIOR FILING DATE: 2000-12-05
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3876
; TYPE: DNA
; ORGANISM: botulinum toxin
US-10-051-952-4
Alignment Scores:
Pred. No.: 7,83e-227 Length: 3876
Score: 2364.00 Matches: 447
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 2
Query Match: 98.87% Indels: 0
DB: 39 Gaps: 0
US-09-910-186a-10 (1-450) x US-10-051-952-4 (1-3876)
```

QY 2 ThrileProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsn 21  
 DB 2527 ACAATACCCCTTTAATATTTTTCATATACACATTAATTCCTTTAATAAGATATAATAT 2586  
 QY 22 GluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr 41  
 DB 2587 GAATATTTCAATAATATTAATGATTCATAAATTTTGGCCTACAAACAGAAAAATACT 2646  
 QY 42 LeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsn 61  
 DB 2647 TTAGTGGATACATCAGGATATATGCGAAGTGTGAGTGAAGAGCGCATGTTGAGCTTAT 2706  
 QY 62 ProIlePhePropheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIle 81  
 DB 2707 CCAATATTTCCATTTGACTTTAAATAGTAGTTTCAGGGAGGATAGAGGTAAGGTATA 2766  
 QY 82 ValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrClnSerPheSerIleSerPhe 101  
 DB 2767 GTACCCAGAAATGAATATTTGATATATCTATGATGAAAGTTTATGACATATGTTT 2826  
 QY 102 TrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerVal 121  
 DB 2827 TGGATTAGATAAATAATGCGTAAGTAATTTTACCTGGATATATCTATAATGATAGTGT 2886  
 QY 122 LysAsnAsnSerGlyTyrSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLys 141  
 DB 2887 AAAATAACTCAGGTGGAGTAGGATAGTATTATTAGTAATTTTATGATTTTACTTTAAA 2946  
 QY 142 GlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaPro 161  
 DB 2947 CAAATGAAGATAGTGAACAAAGTATAAATTTAGTTATGATATACAAATATGCTCCT 3006  
 QY 162 GlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181  
 DB 3007 GGATAGAATAAATGCTTTTGTGAACGTGTACTAACAATAATGATGGAATAATGAGATT 3066  
 QY 182 TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPhe 201  
 DB 3067 TATATAATGCAAAATTAATAGATACTATAAAGTTAAGAACTACTGGAATTAATTT 3126  
 QY 202 SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAsp 221  
 DB 3127 AGCAAACTATAAATGTTGAATTAATAAATTCAGATACCGGTTGATTTACTTCAGAT 3186  
 QY 222 SerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGly 241  
 DB 3187 TCTGATACATCAATATGCGATAGAGATTTTATATATTTCTTAAGAAATAGATGTT 3246  
 QY 242 LysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrp 261  
 DB 3247 AAGAGATATTAATATTAATTAATAGCTTCAATATATCTAATCTTTGTAAGAGATTATGG 3306  
 QY 262 GlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArg 281  
 DB 3307 GGAATGATTAAGATATATAAAGATATATATGTTTAAATAGATATTTTAATATAGA 3366  
 QY 282 TyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPhe 301  
 DB 3367 TATATGTTGCGAACTCAGCAAAATGTTTAAATACAGTAGAATAATAATGACTTC 3426  
 QY 302 AsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgVal 321  
 DB 3427 AATGAAGATATAAATTAATAAAGAAATCAGAGAAATCAAAATGATACATGAGTA 3486  
 QY 322 ArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPhe 341  
 DB 3487 CGAGGAGGAGATATTTATATTTTATGATGACATTAATAACAAGCATATAATTTGTT 3546  
 QY 342 MetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGly 361  
 DB 3547 ATGAAGAATGAACATGATGTCAGATAATCATAGTACTGAAGATATATATGCTATAGT 3606

QY 362 LeuArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleGlnProMetAsn 381  
 DB 3607 TTAAGACAACAACAAGGATATAAATGATATATATATATATATATATATATATATAT 3666  
 QY 382 AsnThrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSer 401  
 DB 3667 AATATCTATTATTAGCATCTCAATATATTAATCAAAATTTTAATGGAGANAATATTTCT 3726  
 QY 402 GlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyLysAspTyrArgHisAsn 421  
 DB 3727 GGAATATGTTCAATAGTACTTATCGTTTAGACTTGGAGTGATTTGGTATAGACAACT 3786  
 QY 422 TyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThr 441  
 DB 3787 TATTGTGCTGCTACTGTGAAGCAAGGAATATATGCTTCATTATTAGAATCAACATCACT 3846  
 QY 442 HisTrpGlyPheValProValSerGlu 450  
 DB 3847 CATGGGGTTTTGTACCTGTAACTGAA 3873

RESULT 12  
 PCT-US97-15394-61  
 ; Sequence 61, Application PC/TUS9715394  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Williams, James A.  
 ; APPLICANT: Thalley, Bruce S.  
 ; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
 ; TITLE OF INVENTION: Botulinum Neurotoxin  
 ; NUMBER OF SEQUENCES: 82  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Medlen & Carroll  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US97/15394  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ingolia, Diane E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: OPND-02304  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 61:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1502 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 108..1493  
 PCT-US97-15394-61

Alignment Scores:  
 Pred. No.: 8.09e-223 Length: 1502  
 Score: 2319.00 Matches: 439  
 Percent Similarity: 98.66% Conservative: 3  
 Best Local Similarity: 97.99% Mismatches: 5  
 Query Match: 96.99% Indels: 1  
 DB: 1 Gaps: 0

US-09-910-186A-10 (1-450) x PCT-US97-15394-61 (1-1502)

QY	4	ProPheAsnIlePheSerTyr_ThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTy	23
DB	152	CCATATCGAAGGTCGTCATATGCGTACGATGCGCTTTATTAAGACATATAATTAATGAATA	211
QY	23	rPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa	43
DB	212	TTTCAATAATATTAATGATTCAAAATTTTGAGCCTCAAAACAGAAAATACTTTAGT	271
QY	43	lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIl	63
DB	272	GGATACATCAGGATATAATGCAGAAAGTGAGTGAAGAGCGGATGTTTCAGCTTAATCCAA	331
QY	63	ePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleValTh	83
DB	332	ATTTCATCTTGACTTTAAATAGGTAGTTCAGGGAGGATAGAGGTAAAGTATTAGTAAC	391
QY	83	rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrpIl	103
DB	392	CCAGAATGAAATATTGTATATAATTCATGTATGAAAGTTTACCATAGTATTTTCGAT	451
QY	103	eArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAs	123
DB	452	TAGAATAAAATAAATGGGTAAAGTAATTCCTCGATATACTATAATTGATAGTGTAAAA	511
QY	123	nAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAs	143
DB	512	TAACTCAGGTGGAGTATAGGTATTATTAGTAAATTTTATTAGTATTACTTTAAACAAAA	571
QY	143	nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTy	163
DB	572	TGAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAAAATATGCTCCTGGATA	631
QY	163	rAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTrpIl	183
DB	632	CAATAAATGGTTTTTGTAACTGTACTAACAAATATGATGGGAATATGAAGATTATAT	691
QY	183	eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLy	203
DB	692	AAATGGAAAAATTAATAGATACATATAAAGTTAAAGAACTAACTGGAAATTAATTTAGCAA	751
QY	203	sThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs	223
DB	752	AACTATAACATTTGAAATAAATAAATCCAGATACCGTTTGATTTACTTCCAGATTCTGA	811
QY	223	pAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs	243
DB	812	TACATCAATATCTGGATAGAGATTTTATATATTTGCTAAAGAAITAGATGTTAAGA	871
QY	243	pIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAs	263
DB	872	TATATAATATTTAATATAGTTCGAATATACTAATGTTGTAAGAAATTTTGGGGAAA	931
QY	263	nAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMe	283
DB	932	TGATTTAAGATATAATAAGAAATATTATATGGTTAATATAGATATTATTAATAGATATAT	991
QY	283	tTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGl	303
DB	992	GTATGCGAACTCACGCAAAATTTGTTTTAATACACGTAGAAAATAATAGACTTCAATGA	1051
QY	303	uGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl	323
DB	1052	AGGATATAAAATTTAATAAAAAGATCATCAGAGGAATAACAATGATCTAGAGTACGAGG	1111
QY	323	yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLy	343
DB	1112	AGGAGATATTTTATATTTTCATATGACAATTTAATAACAAACCATATAATTTGTTTATGAA	1171
QY	343	sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr	363
DB	1172	GAATGAAACTATGTATGCAGATAATCATAGTACTCAAGATATATATGCTATAGTGTTAAG	1231

Qy 363 gGluGlnThrLysAspIleasAspAsnIlelePheGlnIleGlnProMetAsnAsnTh 383  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1232 AGAACAAACAAGGATATAAATCATTAATTATTTCAAATAACCAACAATGAATAATC 1291  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 383 rTyTyTyTrAlaserGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIl 403  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1292 TTATTATTAGGCATCTCAAAATATTANAATCAAAATTTTATGGAGAAAATNTTTCUGNAAT 1351  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 403 eCySserIleclgThTyArGpHeArgLeugLyGlyAspTrpTyArGHisAsnTyrlE 423  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1352 ATGTTCATATAGGTACTTATCGTTTTAGACTTGAGGTGATTGGTAGACACAATTAATT 1411  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 423 uValProthrValIySGlncGlyAsnTyrrAlaserLeuLeuGluSerThrHisr 443  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1412 GGTGCCCTACTGTGAACCAAGAAATATGCTTCATATTAGAATCAACATCAACTCAT 1471  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Qy 443 pGlyPheValproValserGlu 450  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 1472 GGGTTTGTACCCTGTAAGTAA 1493  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
  
RESULT 13  
US-08-704-159-61  
; Sequence 61, Application US/08704159  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thalley, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/704,159  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 61:

Qy 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyr 23  
Db 152 CQATATCGAAGTCTCATATAGCGTACATGGCTTTATTAAAGATATATTAATGAATA 211  
Qy 23 rPheAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43  
Db 212 TTTCAATAATATTAATGATTCACAAATTTGAGCCTACAAAACAGAAAAATACTTACT 271  
Qy 43 IAspThrSerGlyTyrAsnAlaGluValSerGluGluLysAspValGlnLeuAsnProL 63  
Db 272 GGATACATCAGATATATATGCAAGAGTGAAGAGGCGATGTTTCAGCTTAATCCAAT 331  
Qy 63 ePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValIleValTh 83  
Db 332 ATTTCCATTTGACTTTAAATAGTAGTTTCAGGGGAGGATAGAGGTAAGTTATAGTAAC 391  
Qy 83 rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheThrprl 103  
Db 392 CCAGATGAAATATATGTAATAATCTATGATGAAGTTTGTAGCATTAGTTTGGAT 451  
Qy 103 eArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSerValLysAs 123  
Db 452 TAGAATAATAATGGTGAAGTAAATTTACCTGGATATCTATATATGATAGTCTTAAGA 511  
Qy 123 rAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAs 143  
Db 512 TAACTCAGGTTGGATATAGTATATATAGTAAATTTTGTATTTTACTTTAAACAAA 571  
Qy 143 nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTyr 163  
Db 572 TCAAGTAGTAGAACAAAGTAAATTTAGTATGATATATCAAAATAGCTCTCGATA 631  
Qy 163 rAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIl 183  
Db 632 CAATAAATGGTTTTTGTAACTGTACTAAATATGATGATGATGATGATGATGATGAT 691  
Qy 183 eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerL 203  
Db 692 AAATGGAAAAATATAGATATATATAAGTTAAAGAACTAACTGGAATTAATTTAGCAA 751  
Qy 203 sThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223  
Db 752 AACTATACAATTTGAATAAATAAATAATCCAGATACCGGTTGATTAATTAATTAATTA 811  
Qy 223 rAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs 243  
Db 812 TAACTCAATATGCGTAAGAGATTTTATATATTTGCTAAAGAAATAGATGGAAGA 871  
Qy 243 pIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAs 263  
Db 872 TATTAATATATTTAATAGCTTGAATATACATAATGTTGTAAGAAATATTTGGGAAA 931  
Qy 263 nAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMe 283  
Db 932 TGATTAAGATATAATAAAGAAATATATATGTTAAATAGATATATTAATTAATAGATAT 991  
Qy 283 tTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGln 303  
Db 992 GTATCGGAACACACCAAAATGTTTTTAATACACGTAGAAATAATATGACTTCATGA 1051  
Qy 303 uGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323  
Db 1052 AGGATATAAATATATAAATAAAGAAATCAGAGGAAATACAAATGATAGTACAGG 1111  
Qy 323 yGlyAspIleLeuTyrPheAspMetThrIleAspAsnLysAlaTyrAsnLeuPheMetLy 343  
Db 1112 AGGAGATATTTATTTTGTATGACAAATTAATAACAAACCATATATTTGTTATGAA 1171  
Qy 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363  
Db 1172 GAATGAAACTATGTATGAGATATATCATAGTACTGCAAGATATATATGCTAGGTTAAG 1231

Qy 363 gGluGluThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsnTh 383  
Db 1232 AGAACAAACAAAGATATAATATGATATATATATATATTTCAATCAACCAATGAATATAC 1291  
Qy 383 rTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIl 403  
Db 1292 TTAUTATTACGATCTCAATATATTAATCAAAATTTAATGAGAAAAATATTTCTGGAAT 1351  
Qy 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyGlyAspTrpTyrArgHisAsnTyrLe 423  
Db 1352 ATGTTCAATAGTACTTATCGTTTACACTGGAGGTATGTTGATAGACACAAATATTT 1411  
Qy 423 uValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSerThrHistr 443  
Db 1412 GGTGCTACTGTGAAGCAAGAAATATGCTTCATATTAGAAATCAACATCAACTCATTC 1471  
Qy 443 pGlyPheValProValSerGlu 450  
Db 1472 GGGTTTTGTACCTGTAAGTAA 1493  
RESULT 14  
US-10-271-012-61  
; Sequence 61, Application US/10271012  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/271.012  
; FILING DATE: 15-Oct-2002  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/704,159  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-02304  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 61:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1502 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 108..1493  
; SEQUENCE DESCRIPTION: SEQ ID NO: 61:  
US-10-271-012-61  
Alignment Scores: 8.09e-223 Length: 1502  
Pred. No.: 2319.00 Matches: 439  
Score: 98.66% Conservative: 3  
Best Local Similarity: 97.99% Mismatches: 5

```
Query Match: 96.99% Indels: 1
DB: 43 Gaps: 0
US-09-910-186a-10 (1-450) x US-10-271-012-61 (1-1502)

QY 4 ProPheAsnIlePheSerTyr-ThrAsnAsnSerLeuLeuLysAspIleLeuAsnGluTy 23
DB 152 CCATATCGAAGGTCGTCATATGCTAGCATGCTTTATTAAAGATATAAATAATGAATA 211
QY 23 rPheAsnAsnIleLeuAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuVa 43
DB 212 TTTCATATATATATGATTCACAAATTTTGGCCCTACAAACAGAAATAATCTTAGT 271
QY 43 lAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGlnLeuAsnProIl 63
DB 272 GGATACATCAGGATATAATCGCAGAAGTCAGTGAAGAAGCGATGTCAGCTTAATCCAAT 331
QY 63 ePheProPheAspPheLysLeuGlySerSerClyclulaspArgGlyLysValIleValTh 83
DB 332 ATTTCCATTGGACTTTAAATTAGGTAGTTCAGGGAGGATAGAGTAAAGTTATAGTAAC 391
QY 83 rGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSerPheTyrIl 103
DB 392 CCAGAATGAATAATTTGATATAATTCATGATGAAAGTTTACCATAGCTTTTGGAT 451
QY 103 eArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleleaspSerValLysAs 123
DB 452 TAGAATAAATAATGCGTAAGTAAGTAATTTACCGGATATATCTAATGATGTTAAAAA 511
QY 123 nAnSerGlyTyrPserIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAs 143
DB 512 TAACTCAGGTGGATATAGGTATATTATTAGTAATTTTATTAGTATTACTTTAAACAAAA 571
QY 143 nGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTy 163
DB 572 TGAAGTAGTAGAACAAAGTATAATTTTATGATGATATCAATATCAATATGCTCTGGATA 631
QY 163 rAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIl 183
DB 632 CAATAATGGTTTTTGTAACTGTACTACAAATATGATGGAATATGAAGATTTATAT 691
QY 183 eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLy 203
DB 692 AAATGGAATAATTAAGATACATATAAAGTTTAAAGAACTAACTGGAATTAATTTAGCAA 751
QY 203 sThrIleThrPheGluLeuAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223
DB 752 AACTATAACATTTGAATAATAATAATAATCCAGATACCGGTTTGATCTTCAGATTCGA 811
QY 223 pAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs 243
DB 812 TAAACATCAATATGTGGATAAGAGATTTTATATATTTCCTAAAGAAATAGATGTAAGA 871
QY 243 pIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValLysAspTyrTrpGlyAs 263
DB 872 TATTAATATATTATTAAATAGCTTGCATATATCAATATGTTGTAAGAAATTTGGGAAA 931
QY 263 nAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMe 283
DB 932 TGATTTAAGATATAATAAAGATATATATGTTAATATAGATATTATTATAAGATATAT 991
QY 283 tTyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGl 303
DB 992 GTATGGGAACCTCAGCAAAATTTGTTTAAATACAGGTAGAATAATATATGACTTCAATGA 1051
QY 303 uGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGl 323
DB 1052 AGGATATAAATTAATAATAAAGATCAGAGAAATACAATGATGACTAGTACAGG 1111
QY 323 yGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLy 343
DB 1112 AGGAGATATTTTATATTTTGATATGACAAATTAATAACAAGCATATATATTTGTTTATGAA 1171

QY 343 sAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuAr 363
DB 1172 GAATGAACACTATGTATGCAGATAATCATAGTACTGAGATATATATGCTATAGGTTAAG 1231
QY 363 gGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPrometAsnAsnTh 383
DB 1232 AGAACAACAAGGATATAAATGATAATATTATATTCAAAATACAACCAATGAATAATAC 1291
QY 383 rTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGlyIl 403
DB 1292 TTATTATACGATCTCAATATTTTAAATCAATTTTATGGAGAAATATTTCTGGAAT 1351
QY 403 eCysSerIleGlyThrTyrArgPheArgLeuGlyAspTyrArgHisAsnTyrLe 423
DB 1352 ARGTTCAATAGTACTTATCGTTTATAGACTTGGAGGTGATGATGATACACAATTTATT 1411
QY 423 uValProThrValLysGlnGlyAspTyrAlaSerLeuLeuGluSerThrSerThrHisTr 443
DB 1412 GGTGCTACTGTGAAGCAGGAATTTATGCTTCATTATTAGTAACAATCACTCACTCATG 1471
QY 443 pGlyPheValProValSerGlu 450
DB 1472 GGGTTTGTACCTGTAAGTGAA 1493

RESULT 15
US-09-611-419A-11
; Sequence 11, Application US/09611419A
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard A.
; APPLICANT: Byrne, Michael P.
; APPLICANT: Middlebrook, John L.
; APPLICANT: Lapenotiere, Hugh
; APPLICANT: Clayton, Michael A.
; APPLICANT: Brown, Douglas R.
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; FILE REFERENCE: A33626 067252.0105
; CURRENT APPLICATION NUMBER: US/09/611.419A
; CURRENT FILING DATE: 2000-07-05
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/146,192
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1374
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on BONTA HC
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)...(1362)
US-09-611-419A-11

Alignment Scores:
Pred. No.: 7, 949-77 Length: 1374
Score: 868.50 Matches: 197
Percent Similarity: 62.06% Conservative: 86
Best Local Similarity: 43.20% Mismatches: 145
Query Match: 36.32% Indels: 28
DB: 23 Gaps: 11
```

US-09-910-186A-10 (1-450) x US-09-611-419A-11 (1-1374)

Qy	2	ThrllepropheasnlellepheSerTythrAsnAsnSerLeuLeuLysAspIleleAsn	21
Db	49	ACCATGCCATTCAACATCTTCCACACCAACTCCTTGTGAAGGACATCATCAAC	108
Qy	22	GlutyrPheAsnAsnleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThr	41
Db	109	GAGTACTTCAACTCCATCAACGACTCCAAGATCTGTCTTGCAGAACAGAAGAGCC	168
Qy	42	LeuValAspThrSerGlyTyrrAsnGluValSerGluGluGlyAspValGlnLeuAsn	61
Db	169	TTGGTGCACACTCCGGTTTCAACGCCGAGGTCAGAGTCGGTGACAGCTCCAGTTGCAC	228
Qy	62	ProilepheProPheAspPheLysLeuGlySerSerGlyGluAspArgGlyLysValle	81
Db	229	ACCATCTACCAACCACTTCAAGTTGTCTCTCCGGTGAC-----AAGATCATC	279
Qy	82	ValThrGlnAsnGluAsnIleValTyrrAsnSerMetTyrrGluSerPheSerIleSerPhe	101
Db	280	GTCAACTTGNACAAACAACATCTTGTACTCCGGCATCTACGAGAACTCCTCTGTCTCCTTC	339
Qy	102	TrpIleArlgleAsnLysTrpValSerAsnLeuProGly---TyThrIleIleAspSer	120
Db	340	TGGATCAAGATCTCCAAAGACTTGAACCACTCCCAACAGAGTAGTACACCATCATCAACTCC	399
Qy	121	ValLysAsnAsnSerGlyTrpSerIleGlyIleLeuSerAsnPheLeuValPheThrLeu	140
Db	400	ATCGAGCAGAACTCCGGTTGGAAGTTGTATCCGTAAACGGTAACATCGAGTGGATCTTG	459
Qy	141	LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrrAspIleSerAsnAla	160
Db	460	CAGAGCTCAACCGTAAGTACAAAGTCCTTGATCTTCGACTACTCCGAGTCTCTGTCCAC	519
Qy	161	ProGlyTyr---AsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMet	179
Db	520	ACCGGTTACACCAACAAGTGGTCTCGTCCACCATCAACAACAACATCATGGTTACATG	579
Qy	180	LysIleTyrrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyLe	199
Db	580	AGTTGTACATCAACCGTAGTTGAAGCAGTCCCAAGAGATCGAGAGCTCGAGAGCTG	639
Qy	200	AsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr	219
Db	640	AAGTGTGACAAAGACCATCGTCTCGTATCGACGAGAGAACATCGAC-----684	
Qy	220	SerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrrIlePheAlaLysGluLeu	239
Db	685	-----GAGAACCAAGATGTGTGGATCCGTGACTTCAACATCTCTCCAAGAGACTG	735
Qy	240	AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyThrAsnValLysAsp	259
Db	736	TCCAACGAGGACATCAACATCGTCTACGAGGTCAGATCCTGAGAGACGTGATCAAGGAC	795
Qy	260	TyrTrpGlyAsnAspLeuArgTyrrAsnLysGluTyrrMetValAsnIleAspTyrrLeu	279
Db	796	TACTGGGTAAACCCCTGAAGTTCGACACCGAGTACTACATCATCAACGACACTACATC	855
Qy	280	AsnArgTyrrMetTyrrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsn	299
Db	856	GACCGTTACATGCCCCGAGGTCACACGTCCTGGTCTGGTCCAGTACCTGACCGGTTC	915
Qy	300	AspPheAsnGluGlyTyrrLysIleIleLysArgIleArgGlyAsnThrAsnAspThr	319
Db	916	AAGCTGTACACCGGTAAACCTATCACCATCAAGTCCGCTCCGACAGAACCCTTACTCC	975
Qy	320	ArgValArgGlyLysAspIleLeuTyrrPheAspMetThrIleAsnAsnLysAlaTyrrAsn	339
Db	976	CGTATCTTGAAACCGGTGACAACTCATCTCGACATGCTGTGTACACTCCCGTAAGTACATG	1035
Qy	340	LeuPheMetLysAsnGluThrMetTyrrAla-----AspAsnHisSerThrGluAsp	356

Db	1036	ATGATCCGTGACACGACACACCATCTACGCCACCCAGGGTGGTGACTGTTCCGACAACTGT	1093
Qy	357	IleTyrAlaIleGlyLeuArgGluInThrLysAspIleAsnASPAsnIleIlePheGln	376
Db	1096	GTCTACGCCCTGAAGCTGAGCATGCCAACCTGGGTAATCTACGGTATCCGTT	1152
Qy	377	IleGlnProMetAsnAsnTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsn	396
Db	1153	ATCAAGACATCTCTCCAAAGAACAGTACTGCTCCAGATCTTC	1209
Qy	397	GlyGluAsnIleSerGlyIleCysSerIle---GlyThrTyrArgPheArgLeuGlyGly	415
Db	1210	---GAGAAACACCATGTGCTGGCGGCACATCTACAAGCCTTGGCGTTTCTCC	1257
Qy	416	AspTrpTyrArgHisAsnTyrIleuValProThrValLysGlnGlyAsnTyrAlaSerLeu	435
Db	1258	-----TTCAAGAACGCCCTACACTCTGTGCGCGTCACC-----AACTACGAGACCAAG	1305
Qy	436	LeuGluSerThrSerThrHisTrpGlyPheVal-ProValSerGlu	450
Db	1306	CTGCTGTCCACCTCCCTCTGTCGAGAGTTCATCTCCCGTGACCCAG	1351

Search completed: November 7, 2002, 20:04:45  
Job time : 3179 secs





```
Db 1120 GTTCCATTGATCAATGCCATTTACAGCCTCTTTAAAGGCCATAATCCTAATATCTAT 1179
Qy 14 SerLeuLeuLysAspIleLeuAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeu 33
Db 1180 ACTTTGTTACAAATGATA-----GAATCAAAAGG-----AGAGAATATGAATATTA 1227
Qy 34 SerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAla-----Glu 51
Db 1228 GTCAGGAAGATTCACAAAGATACTTTGAAGAATAAGAAAGTTTATACAAGCAAGAGAA 1287
Qy 52 ValSerGlnGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLysLeuGly 71
Db 1288 ATTGAGCAGTATGGTGTGTA----- 1308
Qy 72 SerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyr--- 90
Db 1309 -----AACGACACATCGATAAATAATAATAATATTTGTTATATAG 1350
Qy 91 -----AsnSerMetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLys 107
Db 1351 CATGCCCGAATAGGAAATACATTCATTTGGTGATAAATATATTTATATATAAATGAG 1410
Qy 108 TrpValSerAsnLeuProGlyTyrThrIleAspSerValLysAsnAsnSerGlyTrp 127
Db 1411 CAGAGTATT-----ATGAGTATATAAAGTCAATTAACATATGTCAAATGAA 1458
Qy 128 SerIleGlyIleIleSerAsnPheLeuValPheThrLeu----- 140
Db 1459 AGCGTAGAGGCTATTCAAAATCAGTAGAGTTTGGCTAAATATGAATATACATCGTTG 1518
Qy 141 -----LysGlnAsnGluAsp-----SerGlnGlnSerIleAsnPhe 152
Db 1519 GATTGTAAGAAAGAGGAAATACCTTTAAGAAATGTTTCAAAATAGCAAAATATCGATG 1578
Qy 153 SerTyrAspIleSerAsnAsnAlaProGlyTyrAsnLysTrpPhePheValThrValThr 172
Db 1579 GTATAC-----GGAGTGTGTGGACAGGAAATCACTTTAATAAACCATAT 1626
Qy 173 AsnAsnMetMetGlyAsnMetLysIleTyrIle-----AsnGlyLysLeuIleAsp 189
Db 1627 TGCATTTTCTATGATAAAGATGTTATCGTCAAGCAAAATFACATAACTGCGAGTAGAC 1686
Qy 180 ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle 209
Db 1687 AATATTAAGAAAA-----ATCAAAATATCTAACATTAATAAACGCTACTATT 1734
Qy 210 AsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIle 229
Db 1735 TCTAAA-----TTTTATATATGATAAAGAAAGATGACTGTGTTAATA 1779
Qy 230 ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn 249
Db 1780 ATCGAT-----GAGGCAGGTACAGTTAGTAAATAAGGACATGAATCGAATTCCTGAA 1830
Qy 250 SerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLys 269
Db 1831 ACAAGCAATTGAAATATTA-----TTAATGTGCGGTAT 1866
Qy 270 GluTyrTrpMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsn----- 286
Db 1867 AATTATCAATCGAATCGATAGATTTTGGAAATTTGGTTCGAAATGCCAAAGATGTTTG 1926
Qy 287 SerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyTyrLys 306
Db 1927 TCAAAAAATAAATC-----AACGAATACTACTATGATGAT----- 1962
Qy 307 IleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIle 326
Db 1963 -----CGAATAAATAATGATGATTTA 1983
Qy 327 LeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThr 346
Db 1984 CTTTACTTTTGGAAATCTGTTAGAAAAAAGAGATAATTA-----AATGAAT 2034
Qy 347 MetTyrAlaAspAsnHisSerThr----- 354
Db 2035 ATTAATGATAAATAATATTCTACAGATTAGTAGAAGATATATTTAATGAATTCATAAAA 2094
Qy 355 -----GluAspIleTyrAlaIleGlyLeuArgGlu--- 364
Db 2095 GATCAAAATATTCTTTGTTTAAATACGATGTATATATGTTATCAATCAATTAATAGA 2154
Qy 365 -----GlnThrLysAspIleAsnAsnIleIlePheGlnIleGlnPro----- 379
Db 2155 TTATTACAAGCAAAATAAATAATGATCTGTAATTTGGGTGTGAAGAATATAAAGTT 2214
Qy 380 -----MetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsn 394
Db 2215 GGTGATCCTATTCTATTATGAACTAATAAGTACTACCACTACTTTT---AATAAT 2271
Qy 395 PheAsnGlyGluAsnIle 400
Db 2272 TTAAGAGGTCAATAATT 2289
RESULT 2
US-10-092-411A-879
; Sequence 879, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/10/092,411A
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 879
; LENGTH: 4008
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-879
Alignment Scores:
Pred. No.: 0.000521 Length: 4008
Score: 121.50 Matches: 97
Percent Similarity: 34.83% Conservative: 74
Best Local Similarity: 19.76% Mismatches: 193
Query Match: 5.08% Indels: 127
DB: 6 Gaps: 23
US-09-910-186A-10 (1-450) x US-10-092-411A-879 (1-4008)
Qy 10 TyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAsp 29
Db 2698 TACCTTAACCGTACTATTCCTT-----ATCAATPAACACGCTACTCAAGTAAATAC 2748
Qy 30 SerLysIleSerLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsn 49
Db 2749 ACGTATGTACTACTTCAAGATGGAACAGTAATACTCCATTA----- 2790
Qy 50 AlaGluValSerGluGluGlyAspValGlnLeuAsnProIlePheProPheAspPheLys 69
Db 2791 -----GGATGGGTAAACATTAAATGATGTGACAACTCAAAAT----- 2826
Qy 70 LeuGlySerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleVal 89
Db 2827 ATCGAAAAACAACACTCAGTCTATAGTAAATATTCAGTAAACCTACAAATAATGTCCTA 2886
Qy 90 TyrAsnSerMetTyr----- 94
Db 2887 TATTCTTGTCTGGGTGCTATAAAACCAACAATTAAGTACCTTAATACGCTAGCTAAT 2946
```

QY	95	GluSerPheSerIleThrPheArgLeuAsnLysTrpValSerAsnLeuProGly	114
Db	2947	CAAGCATTTAATCTCCAAAGCTGTTCACGTTGGTAAA-----GAT	2988
QY	115	TyrThrIleIleAspSerValLysAsnSerGlyTyrTrpSerIleGlyIleIleSerAsn	134
Db	2989	TTATACTATACCGGTACAGTCAATACAGACGAGTG-----ATTGCGCTAAG	3039
QY	135	PheLeuValPheThrLeuLysGlnAsnGluAspSerGluInSerIleAsnPheSerTyr	154
Db	3040	GATTTAATC-----CAAAACAGTACTGACGCTCAATCAACCACTATAACTAT	3087
QY	155	Asp---IleSerAsnAlaProGlyTyrAsnLysTrpPhePreValThrValThrAsn	173
Db	3088	ACTTTGTTTCAATTAATAGTAAAGTTAT-----TCTATATGGATCCCAACAAA	3138
QY	174	AsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysVal	193
Db	3139	GCAAACCGATATCTTTAAACCAATATTATGA---CAAACITTCACAGTCAATTAAGCAA	3195
QY	194	LysGluLeuThrGlyIleAsnPhe-----	201
Db	3196	AAAAATTAATGGCTTAATGGCTACTGTCACCTTTAGACGGTAATAATGTTGG	3255
QY	202	SerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThr	219
Db	3256	ATAAATCAACTGACTTAGTAAAGGAAAAATTAATATGCATATACITGGAATGACITTA	3315
QY	220	SerAspSerAsnIleAsn-----	226
Db	3316	AATAACGCGATAAATCACTCTGCTTAAATATAAACACCAAGTACRAAAATGAGCCT	3375
QY	227	MetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn	245
Db	3376	TTGAANTGGTCAATGCTAATATATCTCAAAATAAA-----ATGCTATGATACACAAG	3429
QY	246	IleLeuPheAsn-----SerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsn	263
Db	3430	CGTTTACTGATTCATCTCTTAAATAT-----CAA	3462
QY	264	AspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMet	283
Db	3463	TTCTTACGTTTATGATCACACCACCAATACTTG---TCAGCACAGCTCTCAATTAATTTA	3519
QY	284	TyrAlaAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsn	299
Db	3520	AAAGGCAAGGTGCTACTTGAACCAACGCGCTGCTATTTAGCCAACGTCACGTAAGTAT	3579
QY	300	AspPheAsnGluGlyTyrLysIleIleLysArgIleArg---GlyAsnThrAsnAsp	318
Db	3580	GGCTAATAGAAATTTATCTATCTACATGCTTTAGTAGAACAAGGTAATGGACITCA	3639
QY	319	ThrArgValArgGlyGlyAspIleLeu-----TyrPheAspMetThrIleAsnAsnLys	336
Db	3640	CAACTTCTTAAGGTGGAGATGTTTTCAAAAGGTAAATTCACAACTAAAACAGGTCACAAA	3699
QY	337	AlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp	356
Db	3700	TACATAATGCTTTGGAATGGTGA-----TTTGACAATTAATGCCATCTGTAGAT	3750
QY	357	IleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln	376
Db	3751	GGTATCAATACGCTAAAAATGCTGGATGGAATCTGCTCTCAAAAGCAATTAATGGTGGC	3810
QY	377	IleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGln-----IlePheLysSerAsn	394
Db	3811	GCTAAATCAATGGAATTCATACGTAAGAAGCAGGACAAATACGCTATATAAAATGCGT	3870
QY	395	PheAsnGlyGluAsnIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGly	414
Db	3871	TGGAAATCCTGCAAA-----CCTGGTACGCATCAATATGCAACTGAT	3912
QY	415	GlyAspTrpTyrArgHisAsn-----	421
Db	3913	ATTAATTTGGCAAAATGTCACAGCCACAAGTATTTAAACAATTTATGATAAAATTTGCTGAA	3972
QY	422	TyrLeuValProThrValLys	428
Db	3973	GTCGTAAGTACTTCCGAATTTCCACATACAAA	4005
RESULT 3			
US-10-092-411A-2477			
; Sequence 2477, Application US/10092411A			
; GENERAL INFORMATION:			
; APPLICANT: Lynn Doucette-Stamm et al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCI			
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: 032796-101			
; CURRENT APPLICATION NUMBER: US/10/092,411A			
; CURRENT FILING DATE: 2002-03-07			
; PRIOR APPLICATION NUMBER: US 09/134,001			
; PRIOR FILING DATE: 1998-08-13			
; PRIOR APPLICATION NUMBER: US 60/064,964			
; PRIOR FILING DATE: 1997-11-08			
; PRIOR APPLICATION NUMBER: US 60/055,779			
; PRIOR FILING DATE: 1997-08-14			
; NUMBER OF SEQ ID NOS: 5676			
; SEQ ID NO 2477			
; LENGTH: 2793			
; TYPE: DNA			
; ORGANISM: Staphylococcus epidermidis			
US-10-092-411A-2477			
Alignment Scores:			
Pred. No.: 0.000526	Length: 2793		
Score: 119.50	Matches: 98		
Percent Similarity: 34.55%	Conservative: 63		
Best Local Similarity: 21.03%	Mismatches: 184		
Query Match: 5.00%	Indels: 121		
DB: 6	Gaps: 23		
US-09-910-186A-10 (1-450) x US-10-092-411A-2477 (1-2793)			
QY	9	SerTyrThrAsn-----AsnSerLeuLeuLysAspIlelle	20
Db	670	ACGTATAAAATATAGATGAAATTTCAAAATCAAGATGAGTTATTAATTTACCAATA	729
QY	21	AsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGln---	39
Db	730	AATGAATGAAATTAAGGTAGACCGTTATCTACACATCTGCCCAACCATCGAGTAAG	789
QY	40	AsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspValGln	59
Db	790	CGTGTAAACCGTA-----AATCAATTAGCGGACAGAACAGGTTCCGAATGTT	834
QY	60	LeuAsnProIlePhePropheAspPhelySerLeuGlySerSerGlyGluAspAggGlyLys	79
Db	835	AATCAATTAATTAAGATTTACTGATCAAAAGTATTACTGAAGATATGATGATGATGGT	894
QY	80	ValIleValThrGlnAsn---GluAsnIleValTyrAsnSerMetTyrGluSerPheSer	98
Db	895	ATATTAAGACATGATGCTGCAAACTTAATCTATGATGTAACCTTTTGAA-----	945
QY			

Qy	150	IleAsnPhSerTyrAspIleSerAsnAsnAlaProGlyTyrAsnLysTyrPhePheVal	159
Db	1069	GCTACAGGACTTATGACACCAACAATAAACAATACCTACACTTTTACAGATTATGTA	1128
Qy	170	ThrValThrAsnAsnMetMetGlyAsnMetLysIle	183
Db	1129	GATAAATATGAAAATATTAAACGCGCACCTTAATTAACATCATACATGATAAATCAAG	1188
Qy	184	-----AsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsn	200
Db	1189	GTTCACAAATATAACACTAAGTAGATGTAGAATATAAGACGGCCCTTTCATCAGTAAT	1248
Qy	201	PheSerThrIleThrPheGluIleAsnLysIleProAsp	218
Db	1249	-----AAACAACTACGGTTGGATATCAAAACCTACGAAATCGGATGCTGACCTT	1302
Qy	219	ThrSerAspSerAspAsnIleAsn	229
Db	1303	CAAGTAGTGTTCACAAACATAGATACGAAAAACCATACAGTTGACGAAACGATTATATT	1362
Qy	230	ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn	249
Db	1363	AACCCCTTCGTTATTTCAGCCAAAGAA	1413
Qy	250	SerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLys	269
Db	1414	GGCGATGACGGTTCACAAATATTTCGACGAT	1467
Qy	270	GluTyrTyrMetValAsnIleAspTyrLeuAsnArg	288
Db	1468	GTTCGAGATAATCAAAATTTTCCAGATAGTACAGAAATTTATGATTACAGGGAATATGAA	1527
Qy	289	GlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIle	308
Db	1528	GATGTC	1554
Qy	309	IleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIle	326
Db	1555	-----TTAGGAATAATAATGACGTGGAATATAATTTGGTAATATAGATTCA	1602
Qy	327	LeuTyrPheAspMetThrIle	341
Db	1603	CCATATATTATAAGATTATTAGTAATATGACCCTAATAAGGACGATTACACGACGATA	1662
Qy	342	MetLysAsnGluThrMetTyrAlaAsp	360
Db	1663	CAGCAAACTGTGACATGCAACGACTATAATAGTACTACTGGTGATTT	1713
Qy	361	GlyLeuArgGluGlnThrLysAspIleAsnAspAsnIlePheGlnIleGln	378
Db	1714	-----AGAACAGCATCTATGATAATACAATTCGTTCTCTCTACAGTTCAGGT	1761
Qy	379	-----PrometAsnAsnThrTyrTyrTyrTyrTyrAlaSerGlnIlePheLys	392
Db	1762	CAAGSACAAGGTGACTTGCCTCTGAAACAACTTATAAAATCGGAGATTACGTATGGGAA	1821
Qy	393	-----SerAsnPheAsnGlyGluAsnIleSerGlyIle	403
Db	1822	GATGTAGTAAAGATGCTATTCAAAATACAAATGATATGAATAAAGCCGCTTAGTAATGTA	1881
Qy	404	CysSerIleGlyThrTyr	409
Db	1892	TTGGTAACTTTGACGTAT	1899

## RESULT 4

```

RESULTS *
US-10-092-411A-1838
; Sequence 1838, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092.411A

```

```
; Sequence 1411, Application US/10240425
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1411
; LENGTH: 2376
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. U02020
US-10-240-425-1411

Alignment Scores:
Pred. No.:          0.000907      Length:       2376
Score:             116.50        Matches:     77
Percent Similarity: 36.17%       Conservative: 59
Best Local Similarity: 20.48%    Mismatches: 133
Query Match:       4.87%         Indels:     107
DB:                6            Gaps:       18

US-09-910-186A-10 (1-450) x US-10-240-425-1411 (1-2376)

Qy   7 IlePheSerTy-----ThrAsnAsnSerLeuLeuLysAsp--- 18
      :::::|||||
Db   124 GTTATTTCCTACTTTGAATGCCGTGCGTAAAGAACACAGACAATACTCCAATTAAAGAAAGGTG 183
Qy   19 -----IleAsnGluTyPhe----- 24
      |||:::::|||||
Db   184 AAATATGAGGAACAGATATTTATGGTTCGCATCATCTTATAAGTACTTAAGAAGGT 243
Qy   25 -----AsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLys 39
      ||| :::::||||| :::::
Db   244 AAAAGTAGTAACCAAGAGAAAATCCAGGAGGCCAAGATGTCTACAAGAACACATTTCCCAA 303
Qy   40 AsnThrLeuValAspThrSerGlyTyTrsAlaGluValSerGluGluGlyAspValGln 59
      ::::: ::::: ::::: ::::: |||||
Db   304 GATGATGCTTTAATGAAAAAGGATGGAAC--TACATTCTTGAGAGATGATGAGGCAT 360
Qy   60 LeuAsnProIlePheProPheAspPheLysLeuGlySerSerGly-----GluAspArg 77
      ||| ||| ::::: ||| ||| |||
Db   361 CTY-----CCAATAGAAATAAAAGCTGTTCCTGAGGGCTTTGTCATTTCCTCCAGA 408
Qy   78 GlyLysValIleValThrGlnAsnGluAsnIleValTyTrsSerMetTyTrsGluSerPhe 97
      ||| ||||: ||| ::::: |::: ||| :::
Db   409 GGAATGTTCTCTTCCACG-----GTGGAACACACACAGATCCAGAGCTGTTAC 453
Qy   98 SerIleSerPheTrpIleArg-----IleAsnLysTrpValSerAsnLeuProGly 114
      ::::: ||||| ::::: |||
Db   454 TGCGTTCAAAATGGATTGAGACATATTCCTTGTCAGTCCCTGG----- 495
Qy   115 TyrThrIleLeuAspSerValLysAsnAsnSerSeryTrpSerIleGlyIleIleSerAsn 134
      ||| ||| ||| |||||
Db   496 TATCCAATC-----ACAGTGGCCACAAAATCTCTAGAGAGACAGAAATAATTTGGCCAAA 549
Qy   135 PheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIle---AsnPheSer 153
      :::||||: ||| ::::: ||| ::::: ||| ::::: |||
Db   550 TATTTCGTAGAAACACTCTGGTAACTTAGATGGTCTCGGAATACAAAGTTACATGATTITGGC 609
```

Qy 154 TyrAsp-----IleSerAsnAsnAlaProGlyTyrAsnLysTrpPhePheValThrVal 171  
 Db 610 TACAGAGAGTCTCTCCCAAGAGACTGCTGGC----- 642  
 Qy 172 ThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIle 191  
 Db 643 -----ATAGAGGATCTGCTCACTTGGTAACTTCACAGAGACACATACAGTA 690  
 Qy 192 LysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLys 211  
 Db 691 -----GCAGAGACTGCTCTAATTAATAAATATTATTGAACGAGAAAGATCCT 735  
 Qy 212 IleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgsp 231  
 Db 736 GTTCAGGCTATTCTGTCAGCAGCAGACACAGATACCATACACCTTGGGGAAAGAC 795  
 Qy 232 -----PheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIle 246  
 Db 796 CATGAAAAAGATGCTTTGAACATATTGTAACACAGATTTCATCAGTCTGCTATCTGTG 855  
 Qy 247 LeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArg 266  
 Db 856 GTACCGATACCTATGACATTTAATGCGTGTGAGAAATATGGGTGAAGATCTAAGA 915  
 Qy 267 TyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsn 286  
 Db 916 CAT-----TTAATAGTATCGAGA 933  
 Qy 287 SerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrLys 306  
 Db 934 AGTACAGAGCA-----CCA 948  
 Qy 307 IleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIle 326  
 Db 949 CTAAATACAGACCTGATCTCGAACCCTTTCGACACTGCTTAAGGTTTGGAGATT 1008  
 Qy 327 Leu-----TyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340  
 Db 1009 TAGGTAAAGATTCTCTGTACTGAGAAGCTCAAGGGTTACAAGTTG 1056

## RESULT 6

US-10-092-411A-405  
 ; Sequence 405, Application US/10092411A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 032796-101  
 ; CURRENT APPLICATION NUMBER: US/10/092,411A  
 ; PRIOR FILING DATE: 2002-03-07  
 ; PRIOR APPLICATION NUMBER: US 09/134,001  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5676  
 ; SEQ ID NO 405  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus epidermidis  
 US-10-092-411A-405

Alignment Scores:  
 Pred. No.: 0.00153 Length: 3456  
 Score: 116.50 Matches: 97  
 Percent Similarity: 34.44% Conservative: 90  
 Best Local Similarity: 17.86% Mismatches: 175  
 Query Match: 4.87% Indels: 181  
 DB: 6 Gaps: 27

US-09-910-186a-10 (1-450) x US-10-092-411A-405 (1-3456)

Qy 3 IleProPheAsnIlePheSerTyrThrAsnAsnSerLeuLeuLysAspIleIleAsnGlu 22  
 Db 214 CTACCAAGTTCTCCAGTACCACACAGCAGCAATACAGCTATCTGTTTCAGTACCAAGGAA 273  
 Qy 23 TyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeu 42  
 Db 274 -----GACGAATCACTGCCAATCTTACGCAACAACAATATACAAA 315  
 Qy 43 Val-----AspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVal 58  
 Db 316 TTAACAAAGTACAAACGATGTTAAACAAATGAATCGACAAATGTCGATAGAAATCGATA 375  
 Qy 59 GlnLeuAsnProIlePhePheProPheAsp---PheLysLeuGlySerSerGlyGluAspArg 77  
 Db 376 GAATTAAT-----TTCCATCAACAATTTAATCTAGGGTTACATTTCAAGATACA 429  
 Qy 78 GlyLysVal-----Lys 83  
 Db 430 CCGGCTGTGACTCTAAGTTGCAAGCATCAGTCAGTACAGACGAATTTATGTATACA 489  
 Qy 84 GlnAsn-----GluAsnIleValTyrAsnSerMetTyrGluSerPheSerIleSer 100  
 Db 490 AGTAACCTGTTGTTTATACAGTAGACTATATCACTGTCATCAATCAGCGTTGAACCTTAAA 549  
 Qy 101 PheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIleAspSer 120  
 Db 550 TTTATGAACGCTAATAATGAA-----GTTGGTATACCAATTATTTTGTGATTAATCAG 603  
 Qy 121 Val-----Lys 122  
 Db 604 ATTGATAACATATGAAGAAGAAATACATTGTGAACCTTTTAATCAAGATCGCAAAA 563  
 Qy 123 AsnAsnSerGlyTrpSerIleGlyIle----- 131  
 Db 664 TCAATCAAGACTGGGATATCAAACTTCAAGATACATTATTACGTTTCAAAAGTTGATCAT 723  
 Qy 132 -----IleSerAsnPheLeuValPheThrLeuLysGlnAsnGlu 144  
 Db 724 CCACAGATGAATGTACAACTTCAAACTTTCTAGTATTTATGATCAACATCGTGA 783  
 Qy 145 AspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlaProGlyTyrAsn 164  
 Db 784 TCAACAGAGACTATGTTAATAGAACAACTCAATTACCGACGCACAA----- 834  
 Qy 155 LysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsn 184  
 Db 835 -----TACATATACATTCAAAATGAATGCAA----- 861  
 Qy 185 GlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThr 204  
 Db 862 --TCTATTCTTGACACCTTCAAAATTAATGA-----GAACAATTCGAGGAAGCA 909  
 Qy 205 ---IleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAsp 223  
 Db 910 TATATTCAATTTCAACAAATACAA-----GAAGTCAGCGCAGAGACACAA 954  
 Qy 224 AsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAsp 243  
 Db 955 TTGCTCAAT-----GAC 966  
 Qy 244 IleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsn 263  
 Db 967 TCTAATCAATATTATTAATTAATTAATAAAGAGCGT-----AAAGATATATTAGATAAT 1020  
 Qy 264 -----AspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle--- 276  
 Db 1021 GCTTATATCATGACGTACGATATCGCGAATCTTACGGATATTATTAGAACATCGGCA 1080  
 Qy 277 ---AspTyr---LeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThr 294  
 Db 1081 ACTGATTTTAAAGTGAATGGATT-----TTAATAAA 1113  
 Qy 295 ArgArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArgGly 314

```

Db 1114 AGGAGAAAAAGAGAGACAA-----ATCAACGACTTAAATGAG 1155
Qy 315 AsnThrAsnAspThrArgValArgGly-----Asp 325
Db 1156 GCGACCACTCAATTGCAAGAGAAAGTAATCAACAAGTACGACACCACTTGGTGAAGAT 1215
Qy 326 IleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGlu 345
Db 1216 ATGTCATTTTAACTAGATCAATAAATAACATCGCTGTAATGAAAAAATACTAATCA 1275
Qy 346 -----ThrMetTyrAlaAspAsnHisSerThrGlu 355
Db 1276 GAATATGACGTCGTCGTCCTCACTTATATCATGAGCTATATCAAACTCAACAGCATAGC 1335
Qy 356 AspIleValAlaIleGlyLeuArgGluInThr--LysAspIleAsnAspAsnIleIle 374
Db 1336 AACACATACGTTTAAACATTTTCAGATGAAGTATTAAGCTTTGAATAAAAATAGAA 1395
Qy 375 PheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsn 394
Db 1396 AATGAGTCAACACCA-----CTATTGAAGAAGCT 1425
Qy 395 PheAsnGlyGluAsnIleSerGlyIleCysSer----- 405
Db 1426 GTCAATCATGATACAAAGTTAATGAATATCGAGTGATCAAAATGAAGATAGGTATGAAT 1485
Qy 406 -----IleGlyThrTyrArgPheArgLeuGlyGlyAspTyrArgHis 420
Db 1486 GATAGATACATTAACATTAACACATTAAGGATTCGCTTACATCCACACATACAAACAT 1545
Qy 421 AsnTyrLeuValProThrValLysGlnGlyAsnTyrAlaSerLeuLeuGluSerThrSer 440
Db 1546 TACTATATC-----CATTTAGACGATCTTTAGATAAATTAATTTGGAAGAACAGAG 1596
Qy 441 ThrHisIleP 443
Db 1597 ACTCATTTT 1605

RESULT 7
US-10-092-411A-2277
; Sequence 2277, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ IDS: 5676
; SEQ ID NO 2277
; LENGTH: 1494
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-2277

Alignment Scores:
Pred. No.: 0.000794 Length: 1494
Score: 114.50 Matches: 90
Percent Similarity: 35.27% Conservative: 56
Best Local Similarity: 21.74% Mismatches: 158
Query Match: 4.79% Indels: 110
Gaps: 21

US-09-910-186A-10 (1-450) x US-10-092-411A-2277 (1-1494)
Qy 10 TyrThrAsnSerLeuLeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAsp 29

```

```

Db 300 TACACGTTTAAACGGCATGAGATGATAAATCAATGAATTCAGTCGACATAGCAACGAT 359
Qy 30 SerLysIleLeuSerLeuGlnAsnArgLys-----Asn 40
Db 360 TGTACACACGTTGCCCTTAGAATTAGAAGGGTTTAATTTATATCATTTTAAATGATTAAAC 419
Qy 41 ThrLeu-----Val-AspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVa 58
Db 420 ACAATTAGATAAACAATGAGTTTCAGGTTACAAA-----TT 455
Qy 58 LglnLeuAsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGl 78
Db 456 TAATTAGAAATCGATTGATCAATTAATTTAGCTGTGGAAGCGGACCTCAAAAGAC 515
Qy 78 YLysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSe 98
Db 516 TGTCCCTCAACAGGTTTAATTAATCATTTAGCCAGTGTAAAGCGTGTAAACAAGTTTAGG 575
Qy 98 rIleSerPheTrpIleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThrIleIl 118
Db 576 ATTCGAA-----CAAAATACTGTGTGGCTTCGGTCTTACCTATATATCATATT-- 624
Qy 118 eAspSerValLysAsnAsnSerGlyTrpSerIle-----GlyIleIleSerAsnPh 135
Db 625 -----TCTGGCTCAGTGTATTTTCCGCGCGCAGTGATGATCAAGGAT 665
Qy 135 eLeuValPheThrLeuLys-----GlnAsnGluAspSerGluGlnSerIleAsnPheSe 153
Db 666 CACTGTGACAGCTGTTTAAAAAGTTTCAAACTGATGATATGTTAAACACAAATAAAG--AC 722
Qy 153 rTyrAspIleSerAsnAsnAla-----ProGlyTyrAsnLysTrpPhePheValThrVa 171
Db 723 TTATCCCAATCACCATATGTCCTTGTCCCAACAGCTTAAAGTGTGTTAATGATGTCGAG 782
Qy 171 lThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly-----Lys 186
Db 783 ATTGACTCAACCATTTTCTTTAGAAAAAATTCGTAGGTGGTCTAAATATCACCACA 842
Qy 186 sLeuIleAsp-----ThrIleLysValLysGluLeuThrGlyIleAsnPheSerly 203
Db 843 ATTAATTGACCAAGCATGATGCTTACCTGCTGATATATATCTTTGGTATGACAGA 902
Qy 203 sThrIleThrPheGluIleAsnLysIlePro-----AspThrGl 216
Db 903 AACTGTCTCTCAGTTTCTACAGCCCTCACCTCAATGCTCAAGAAGCGTTTCGATACT-- 960
Qy 216 yLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIlePheAl 236
Db 961 ---GTTGGAAAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1016
Qy 236 aLys---GluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGluTyrThrAs 255
Db 1017 ACATGGAGAGTTTATTAATTAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1073
Qy 255 nValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValas 275
Db 1074 ATATTAAAAAGACACATTTGATATGAT---GGGTATTTTCAAACTGGAGATATAGCTGA 1130
Qy 275 nIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrAr 295
Db 1131 AATAGAT----- 1137
Qy 295 gArgAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAs 315
Db 1138 -----GATGAAGGTTTACGTCAATAATATATATGATCGG----- 1167
Qy 315 nThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAs 335
Db 1168 -CGCAAGAGATTGATTATAGTGGTGAAGAGATATTTAT----- 1206
Qy 335 nLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGl 355

```

Db	3461	ACCAACATCATCTCTAATACTAAACCTTAAACAAACCAACAAACAAAAAATTTCCAA	3400
QY	96	SerPheSerIleSerPheThrIleArgIleAsnLysTyrpValSerAsnLeuProGlyTyr	115
Db	3401	CCAAATATCCCTA-----ATAAACATCATTAACGAAAAATACATAAATAATACGCAA	3351
QY	116	ThrIleAspSerValIysAsnAsnSerGlyTyrTpSerIleGlyIleIleSerAsnPhe	135
Db	3350	ACTAAATCCACACACATCATCAAAACCTTATCAACCCACATCAAAATTAACCTCATCC	3294
QY	136	LeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAsp	155
Db	3293	-----CTAAATATACAAACTAATTC-AACATACACAAATCAATATAAACAATACATC	3241
QY	156	IleSer-----AsnAsnAlaProGlyTyr	163
Db	3240	GTAAACAAACCAATCAACAAAAACCGTAATATTACTCAATAAATACAAAAAACCTTA	3181
QY	164	AsnLys-----TrpPheValThrValThrAsnAsnMetMetGlyAsnMetIysIle	181
Db	3180	AATAAAATCAACACCCCTTCATATAAACTCAATAAATAAATAAATAAATAAACAAG	3121
QY	182	TyrIleAsnGlyIlystIleAspThrIleLysValIysGluLeuThrGly-----Ile	199
Db	3120	TATCTC-----AAATAATAAATAAACTATTATTATACAAACCCACCAACCAATAACAT	3067
QY	200	AsnPheSerLysThrIleThrPheGluIle-----Asn-LysIleProAspThrGlyLe	217
Db	3066	AATAACAAAAACTTAAAAACGTCCTCTTAAATAACTAACACAAACAAATATACCTCT	3007
QY	217	uIleThrSerAspSerAspAsnIle-----AsnMetThrIleArgAs	231
Db	3006	CACACTCTCTATTCAACATTAATATAAANAATCTTAACCAACATCAACAAAAAANA	2947
QY	231	pPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLe	251
Db	2946	AAATAAAATATATTCAAATAAA-----AAAAAATACTCAAAATTAATATCTA	2901
QY	251	uGlnTyrThrAsnValIysAspTyrTyrGlyAsnAspLeu-----	265
Db	2900	-----TTTACAATAACATAATTATATTTAAAAATCCCAATCTCAACCCAAAATCT	2845
QY	266	-----ArgTyrAsnLys-----GluTyrTyrMetValAsnIleAspTyrLe	279
Db	2844	CCTTAAACTAATCAACAACCTTCAACAAATCTCAAAATTCACAAATCAATAAANAATC	2785
QY	279	uAsnArgTyrMetTyrAlaAsnSerArgGlnIle-----ValPheAsn-----	293
Db	2784	ACAAAATTCCTATACACAATTAATAAACACAAAAACCAATCATAAATAACTCCC	2725
QY	294	-----ThraGArgAsnAs	298
Db	2724	ATTACAGATTACTACAAAAAATAAATACTAATAATCCAAATCAAAAAAATAAATA	2655
QY	298	nAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAs	318
Db	2664	AAACCTATTCAAAAAAATACTACAAACCACTACTCAAAAAATAAAAAAACAATAA	2605
QY	318	pThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTy	338
Db	2604	-----ATAAAAAACATTCCTACTCATTAATAAANA-----	2572
QY	338	rAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp-Ile	358
Db	2571	-AATCAATATCATAAAAAT-----AACCATACTACCCAAAAATAATTT	2531
QY	358	yAlaIleGly-----LeuArgGluGlnThrIleAspIleAsnA	371
Db	2530	ATAAATTCAAATACTATCCCATCAAACTACTACTACTTTTTTTCACAAATTAANA	2471
QY	371	spAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSer-----	388
Db	2470	ACTACTTTAAATTCATATAAAACCAAAAAAATACTTATATACCAAAACATCTCAAC	2411

QY	136	LeuValPheThrLeuLysGlnAsnGluaspSerGluGlnSerIleAsnPheSerTyrAsp	155
DB	3293	-----CTAAATACAAAACTAAATTC-AACATACAAATCAATAAACAATACTATTC	3241
QY	156	IleSer-----	163
DB	3240	GTAACAAACCAATCACAAAACCCAGCGTAATTATCTCAATAAATACAAAAAACCCFTA	3181
QY	164	AsnLys-----TrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle	181
DB	3180	AATAAAATTCAACACCCCTTCATCTATAAACTCTCAATAAATACTAAATATTATAAATAACG	3121
QY	182	TyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGly-----Ile	199
DB	3120	TATCTC-----AAATATATAAACTATTATACAAACCCACCAACCAATAACATACTA	3067
QY	200	AsnPheSerLysThrIleThrPheGluIle-----Asn-LysIleProAspThrGlyLe	217
DB	3066	AATAACAAAAAATAAAACGTTCCCTTTAAAAAATAACACAAAAACAATATATCTCTCT	3007
QY	217	uIleThrSerAspSerAspAsnIle-----AsnMetTyrPileArgAs	231
DB	3006	CACACTCTCTATTACACATATATATAAAATTTCTACCAAAACATCAACAAAAAANA	2947
QY	231	pPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLe	251
DB	2946	AAATAAAATATATTCAATAAATA-----AAAAAAAATACTCAAATTTATATCTA-	2901
QY	251	uGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu-----	265
DB	2900	-----TTTACAAATACATATATATATTTAAAAAATCCCATTTCTCAACCCAAATCT	2845
QY	266	-----ArgTyrAsnLys-----GluTyrTyrMetValAsnIleAspTyrLe	279
DB	2844	CCITAAACATAATCAACAACCTCAACAAAAATCTCAAATTACAAATCAATAAATAAATC	2785
QY	279	uAsnArgTyrMetTyrAlaAsnSerArgGlnIle-----ValPheAsn-----	293
DB	2784	ACAAAATTCCTATACACAAATATAAACAACAACAAAAACCAATCATATAATAACTCC	2725
QY	294	-----	298
DB	2724	ATTCAGATTACTACAAAAAATAAATACTAAAAATCCAACTTACAAAAAATAAATA	2665
QY	298	nAsnAspPheAsnGluGlyTyrIleIleIleLysArgIleArgGlyAsnThrAsnAs	318
DB	2664	AAACCTTACAAAATAAATACTACAAACCATCTCTCAAAAAATAAAAAAACACAAATAA	2605
QY	318	pThrArgValArgGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTy	338
DB	2604	-----ATAAAAAACATTCCTACTCTCATATAATAAATA-----	2572
QY	338	rAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp-Ile	358
DB	2571	-AATCAATATCATAAAAAT-----AACCATACTACCCAAAAATAATTT	2531
QY	358	rAlaIleGly-----LeuArgGluGlnThrLysAspIleAsnA	371
DB	2530	ATAAATTCAAATCTATCCCATCAAACTACTACTACATCTTTTTCAAAATATAAANA	2471
QY	371	spAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSer-----	388
DB	2470	ACTACTTTAAATTCATATAAACCACAAAAAATAACTTATATACCAAAACATCTCAAC	2411

[illegible][illegible][illegible]

	Qy	266	-----ArgTyrAsnLys---GluTyrrMetValasnIleAspTyrLe	279
			:::     :::	
Db	2844	CCTAAACTAATCACAACACTTCAGCAAAATCTCAAATTACAAAATCATATAAANAATC	2785	
Qy	279	uasnArGtyrMetTyrAlaasnsSerArgGlnile-----ValPheAsn-----	293	
		:::     :::	:::	
Db	2784	ACAAAATTCCTATACACAATATAATAACAACAAAAAACCAATCATAAATAAACTCCC	2725	
Qy	294	-----ThraGArgasnAs	298	
		:::	:::	
Db	2724	ATTACAGATTACTACAAAAAATAAATAFACCTAAAAATCCAAATTCANNAATATAAA	2655	
Qy	298	nAsnAspPheAsnGluGlyTyrLysIleIleLysArgIeArgGlyAsnThrAsnAs	318	
		:::     :::	:::	
Db	2664	AAACCTATTCAAAAAAATACTACAAACCCTACTCTAAAAAATAAAAAAACAACAATAA	2605	
Qy	318	pThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTy	338	
		:::     :::	:::	
Db	2604	-----ATAAAAAACATTCCTACTCATATAATAAAAA	2572	
Qy	338	rAsnLeuPheMetLysAsnGluThrMetTyrAlaaspAsnHisserThrGluAsp-Ilet	358	
		:::     :::	:::	
Db	2571	-AATCAATATCATAAAAAT-----AACCATACTACCCAAAAATAATTT	2531	
Qy	358	yralalIleGly-----LeuArgGluGlnThrLysAspIleAsnA	371	
		:::	:::	
Db	2530	ATAAATTCAAATACTATCCCATCAAACTACTACTACTTTTTTCANNAATATAAAAAA	2471	
Qy	371	sPasnIleIlePheGlnIleGlnPrometAsnAsnThrTyrTyrTralaSer-----	388	
		:::     :::	:::	
Db	2470	ACTACTTTAAATTCATATAAAACCAAAAAAATACTATATACCAAAACATCTCTAAC	2411	

Qy	266	-----ArgTyrAsnLys--	-GluTyrTyrMetValAsnIleAspTyrLe	276
Db	2844	CCTTAAACTAATCAACAACCTTCACACAAATCTCAAAATTACAAATTCATATATAAAAAATC		2785
Qy	279	uAsnArgTyrMetTyrAlaAsnSerArgGlnIle	-----ValPheAsn-----	293
Db	2784	ACAAAAATTCCTATACACAAATATATAAACACAAAAACCAAAATCATAAATAAATCCTCC		2725
Qy	294	-----	-----ThrArgArgAsnAs	298
Db	2724	ATTACGAGTACTACAAAAAAATAAAATACCTATAAAATCCAACTTACAAAAAATATATAA		2665
Qy	298	nAsnAspPheAsnGluGlyTyrTyrIleIleLysArgIleArgGlyAsnThrAsnAs	318	
Db	2664	AAACCTATTCAAAAAAATACAAACCACTACTCAAAAAATATAAAAAAACACACAATAA	2605	
Qy	318	pThra-gValArgGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyr	338	
Db	2604	-----ATAAAAAACATCCCATACTACTATAATAAAAA-----	2572	
Qy	338	rAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp-IleT	358	
Db	2571	-AATCAATATCATATAAAAT-	-----AACCATATACCCAAAAATAATTT	2531
Qy	358	YrAlaIleGly-	-----LeuArgGluGlnThrLysAspIleAsnA	371
Db	2530	ATAAAATCAATACTATCCCATCAAACTACTACTACITTTTTTCACAAAATATAAAAAA	2471	
Qy	371	sPAsnIleIlePheGlnIleGlnPrometAsnThrTyrTyrTyrAlaSer	-----	388
Db	2470	ACTACTTTAAATTCATATAAACACAAAAAAACTATATATACCAAAACATCTCTAAAC	2411	

Qy	294	-----ThrArgArgAsnAs	298
Db	2724	ATTACGGATTACTACAAAAAATAAAATACCTAAAAATCCAACTTACAAAAAATATAAA	26655
Qy	298	nAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsnAs	318
Db	2664	AAACCTATCAAAAAAATACTAAACCACTACTCAAAAAATAAAAAAAACACACAATAA	2605
Qy	318	pThrA-gValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTy	338
Db	2604	-----ATAAAAAACATCCATCACTACATAATAAAAA-----	2572
Qy	338	rAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAsp-IleT	358
Db	2571	-AATCAATATCATAAAAAT-----AACCATATACCCAAAAATAATTT	2531
Qy	358	YrAlaIleGly-----LeuArgGluGlnThrLysAspIleAsnA	371
Db	2530	ATAAATCAATACTATCCCATCAAACTACTACTACITTTTTTCACAAAATATAAAAAA	2471
Qy	371	sPAsnIleIlePheGlnIleGlnProMetAsnAsnThrTyrTyrAlaSer-----	388
Db	2470	ACTACTTTAAATTCATATAAAACCAAAAAAAACTATATATACCAAAAACATCTCTAAC	2411

[illegible][illegible][illegible]

550	LeuArgGluGlnThrLysAspIleAsnA	371	spasnllellepHeGlnProMetAsnThrTyrTyrTyrAlaser	388
Db	:::		:::     :::	
2571	-ATCATATCATATAAAT	2530	ATAAATCAATACTATCCCATCAAACTACTACTACTACTTTTTCACAAATATAAAAAA	2471
	:::			
358	yrAlaileGly	371	LeuArgGluGlnThrLysAspIleAsnA	371
Qy				
2530	ATAAATCAATACTATCCCATCAAACTACTACTACTACTTTTTCACAAATATAAAAAA	2471		
Db				
371	spasnllellepHeGlnProMetAsnThrTyrTyrTyrAlaser	388		
Qy	:::     :::			
2470	ACTACTTTAAATTCATATAAACCAAAAAAAACTATATATACCAAAACATCTCTAAAC	2411		
Db				

Db	2530	ATAAATTCATATCCCATCAAACTACTACTACTCTTTTTCACAAATATAAAAAA	2471
Qy	371	spasnllellePheGlnIleGlnProMetAsnAsnThrTyrTyrTyrAlaSer-----	388
Db	2470	ACTACTTTAAATTCATATAAACCAAAAAAATCATATATATACCAAAACATCTCTAAC	2411

bp 2470 ACTGCTTAATTCATATAAAACCAAAAAAACAATTATATACCAAAAAAATCCTAAC 2411



```
Qy 389 -----GlnIlePheLysSerAsnPheA 396
Db 2410 AAAAAAACAAACTAAAAACATCACTACTACCTAACTTCAAACTATATACAAACTATA 2351
Qy 396 snGlyGluAsn 399
Db 2350 ATAACCAAAAC 2340

RESULT 9
US-10-240-485-79/c
; Sequence 79, Application US/10240485
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 79
; LENGTH: 6713
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-79

Alignment Scores:
Pred. No.: 0.0158 Length: 6713
Score: 111.00 Matches: 102
Percent Similarity: 33.21% Conservative: 76
Best Local Similarity: 19.03% Mismatches: 166
Query Match: 4.64% Indels: 195
DB: 6 Gaps: 21

US-09-910-186A-10 (1-450) x US-10-240-485-79 (1-6713)
Qy 4 ProPheAsnIlePheSerTyr-----ThrAsnAsnSer 14
Db 3084 CETHATTAACCTTTTCTCTACTCTTCTATCTAACCACTCTAAATTTCCCTGTAATACTCC 3025
Qy 15 LeuLeuLysAspIleIleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSer 34
Db 3024 CRT-----CCCTCCGCCATAAAT 3007
Qy 35 LeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGlu 54
Db 3006 AGGAAACCCCAAACTATATTTCTTAAACACTCTTTCAATTCCTCTTCATCTCCCTCC 2950
Qy 55 GluGlyAspValGlnLeuAsnProIlePheProPheAspPheLysLeuGly-----71
Db 2949 -----TCITTAATAAAATCTAAC-CCAATTCCTATATCTTTAAATACCATCTAA 2903
Qy 72 -----SerSerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsn 87
Db 2902 ATATAAATCATTAACAAATCAAAACGATATACTTATTCATATAAATACTAAA 2843
Qy 88 IleValTyrAsnSerMetTyrGluSerPheSerIleSerPheIleArgIleAsn---106
Db 2842 ATAACATTTATATAAATTCACATCTCTTTATATAAAATCTCCACCAAAATATATA 2783
```

```
Qy 107 -----LysTyrpValSerAsnLeuproglyTyrThrIleIle 118
Db 2782 AAAAAACATACCTCAAAATATAAACAACATATATAACAACCTCAATCAATCATCATATA 2723
Qy 119 AspSerValLysAsnAsnSerGlyTyrpSer-----128
Db 2722 CT-AAACACAAAAAAATAAAACCTTTCTCCAAAAACTAAAAACAACAAAAATACCC 2664
Qy 128 -----128
Db 2663 CTCCTACCACTATTATTCTACGTAATACAAAAATCTTAACCAAACTATCAACAAAAA 2604
Qy 129 -----IleGlyIleIleSerAsn 134
Db 2603 AAAAAAATAATATCCAAATATAAAAAAATAAATCAAAATTTAATTTATTTCACAA 2544
Qy 135 PheLeuValPheThrLeuLysGlnAsnGluAspSerGlu-----GlnSerIle 150
Db 2543 TACATAATCTTATACCTTAAAAAATCTAAAAATCCACCTAAAAAACTATAATCTAAT 2484
Qy 151 AsnPheSerTyrAspIleSerAsnAlaProGlyTyrAsnLysTyrPhePheValThr 170
Db 2483 CAATAAATTACAAATACAAATCAACATACAAATCAATTAAC---ATTATATAACA 2427
Qy 171 ValThrAsnAsnMetMetGlyAsnMetLys-----IleTyrIleAsnGlyLysLeu 187
Db 2426 ACAATAACAATCTAATAAAAAATCAAAAAACAATTCCTTTTACAATAACACAAAAAT 2367
Qy 188 IleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPhe 207
Db 2366 ATCAATACCTAATAA-----ATCAATTAACATAAAAAATAA---2328
Qy 208 GluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMet 227
Db 2328 -----2328
Qy 228 TrpIleArgAspPheTyrIlePheAlaLysGluLeu-----239
Db 2327 -----TAICTATCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2283
Qy 240 -----AspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVal 256
Db 2282 TCGAAAAAACACAAAAACCAAAATATATATACCTCGTAATAAACTA-----2238
Qy 257 ValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle 276
Db 2237 -----AAAAAATAATATATATAAATAA 2214
Qy 277 AspTyrLeuAsnArgTyrMetTyrAla-AsnSerArgGlnIleValPheAsnThrArgAr 296
Db 2213 CAATAC-----TACCTAAACAAATTTACAAATTCATACAAATTCCTATCA 2169
Qy 296 GAsnAsnAsnAspPheAsnGluGlyTyrLysIleIleIleLysArgIleArgGlyAsnTh 316
Db 2168 AAATACAAATAACCATCTTCACGAAACAAAAAATTCCTAAATTTATAAAACCCAC---2111
Qy 316 rAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLy 336
Db 2110 -----AAAAACCTTAATAACCAACAACTCTTAACAAAAAACAACAA 2067
Qy 336 sAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAs 356
Db 2066 AACTAAAAACATCATCAGTACTACTTCAAAATTTACTA-CAAAACTATATAAATCAACAAA 2008
Qy 356 pIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsn-----372
Db 2007 CAACAAATATTATACATAAACAACAAACATACCAATAAACAAATAAATAAATCCAC 1948
Qy 373 -----IleIlePheGlnIleGlnPrometAsnAsnThrTyrTyrTyrAlaSerGlnIlePh 391
Db 1947 ATATAAATCCATACACTTACAAACCACTAATCTATACAAACTATATAAATAAATACATAAT 1888
```

QY 391 elysSerAsp-----PheAsnGly-----GluAsnIleSerGlyIleCysSerI1 406  
 Db 1887 AAAAAAACAATCTCTTAATAATAATAATAATAATAATAATAATAATAATAATA 1843  
 QY 406 eGlyThrTyr-----ArgPheArgLeuGlyGlyAspIlePheArgHisAsn----- 421  
 Db 1842 AACTACATACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1783  
 QY 422 -----TyrLeuValProThrValLysGlnGlyAsnTyr 432  
 Db 1782 TACCCTTCCT 1737

RESULT 10  
 US-10-240-453-25/c  
 ; Sequence 25, Application US/10240453  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OLEK, Alexander  
 ; APPLICANT: PIEPENBROCK, Christian  
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
 ; TITLE OF INVENTION: Transcription  
 ; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated  
 ; TITLE OF INVENTION: With DNA Transcription  
 ; FILE REFERENCE: 5013.1009  
 ; CURRENT APPLICATION NUMBER: US/10/240,453  
 ; CURRENT FILING DATE: 2002-10-02  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/039973  
 ; PRIOR FILING DATE: 2001-04-06  
 ; PRIOR APPLICATION NUMBER: DE 10019058.8  
 ; PRIOR FILING DATE: 2000-04-06  
 ; PRIOR APPLICATION NUMBER: DE 10019173.8  
 ; PRIOR FILING DATE: 2000-04-07  
 ; PRIOR APPLICATION NUMBER: DE 10032529.7  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: DE 10043826.1  
 ; PRIOR FILING DATE: 2000-09-01  
 ; NUMBER OF SEQ ID NOS: 350  
 ; SEQ ID NO 25  
 ; LENGTH: 5979  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 US-10-240-453-25

Alignment Scores:  
 Pred. No.: 0.0256 Length: 5979  
 Score: 108.50 Matches: 104  
 Percent Similarity: 37.03% Conservative: 83  
 Best Local Similarity: 20.59% Mismatches: 168  
 Query Watch: 4.54% Indels: 151  
 DB: 6 Gaps: 22

US-09-910-186A-10 (1-450) x US-10-240-453-25 (1-5979)  
 QY 7 IlePheSerTyrThrAsnAsnSerLeuLeuLysAsp-----IleLeuAsnGluTyrPhe 24  
 Db 4240 ATATACACCCATTTCACCAATATATTATAAATAAATAAATAAATAAATAAATAAATA 4181  
 QY 25 AsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAsp 44  
 Db 4180 AATAATATCAAAACAATAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4121  
 QY 45 -ThrSerGlyTyrAsnAlaGluValSerGluGlu-----GlyAspVa 58  
 Db 4120 TACTAATTCACCAAAACAATAATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4061  
 QY 58 lGlnLeuAsnProfile---PheProPhe-----AspPheLeuLysGlySerSerG1 74  
 Db 4060 CCAAAAATCCATCAATCACT 4001  
 QY 74 yGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyr----- 90

Db 4000 TCACTAAAAAATCCACTTTAAATTTTACACCAAAACAAAATCAACAATATCAATCTAAAT 3941  
 QY 91 -AsnSerMetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrpValse 110  
 Db 3940 AAATCCATTAATCCAAACATATTTCAATTCATCTACACACAAAATAAATTCGTAA 3881  
 QY 110 r-----AsnLeuProGlyTyrThr-- 116  
 Db 3880 CAAAAACAATATCCAAAATTCATTTTAAATTTTAAACAAAACCAACCACTCATCAAA 3821  
 QY 116 ----- 116  
 Db 3820 TCAATTAACAAATAATCTCGTATATTAACCAATAATATATATATATATATATTAACATA 3761  
 QY 117 -IleIleAspSerValLysAsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLe 136  
 Db 3760 AATCACAACAATAATAATAACACAAAACAATTTAACCTA-----AACTATTT 3713  
 QY 136 uValPheThrLeuLysGlnAsnGluAspSerGluGln----- 148  
 Db 3712 ATTCTTCTCATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 3653  
 QY 149 -----SerIleAsnPheSe 153  
 Db 3652 ACCTACCACCTTTATAATAATTTACCAAACTAATATCAATTAATTAATAATAAATAAATA 3593  
 QY 153 rTyrAspIleSerAsnAsnAla---ProGlyTyrAsnLysTrpPhePheValThrValTh 172  
 Db 3592 CTPAACATTTAAATAATCAACGTAATAATCTATAACAAAATACAAATTCATTTCTTAA 3533  
 QY 172 rAsnAsnMetMetGlyAsn-----MetLysIleTyrI1 183  
 Db 3532 ATCCCAATATACAAAATAAAGATACAAATAATTAATCTACTTCAAAATATATA 3473  
 QY 183 eAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLy 203  
 Db 3472 AAAT---AAATACACACATAATTTCAAATTTTAAATAAT-----ATAAATCTCTCAA 3422  
 QY 203 sThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSerAspSerAs 223  
 Db 3421 AACCTTAACCCATAAATCAAC-----ACTTCTCTCTCTAA 3386  
 QY 223 pAsnIleAsnMetTrpIleArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAs 243  
 Db 3385 TAAT-----AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3362  
 QY 243 pIleAsnIleLeuPheAsnSerLeu---GlnTyrThrAsnValValLysAspTyrTrpG1 262  
 Db 3361 TTTAATATTCATTAATAATAATTTCTATAACAAATAATATTTATATATTAATAAATAAATA 3302  
 QY 262 yAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTy 282  
 Db 3301 AAAC-----CATTTCAATCAACACACAAAATAATTCAAAATAATATATATATATATAT 3248  
 QY 282 rMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsnAsnAspPheAs 302  
 Db 3247 TAAATTTTCGAACACACAAACATTTTTCATTAATTCATTAATTCATTAATTCATTAATTA 3188  
 QY 302 nGluGly-----TyrLysIleIleIleLysArgIle-ArgGlyAsnThrAsnA 318  
 Db 3187 ATCAAAATCTTATTTATACCAATTAATTTTAAATAATATTTTCAATAAATAAATAAATA 3128  
 QY 318 sPThrArgValArgGlyGlyAspIleLeu----- 327  
 Db 3127 AACAATATTTTAAACCAATAACCTTAAACACACACTTTTAAATAAATAAATAAATAAATA 3068  
 QY 328 -----TyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheM 342  
 Db 3067 CATAAACCAAAATAAATAATTTTCGATTACCAATTTTCCAAATTTACCA---ACTTTATCA 3011  
 QY 342 eTylAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyL 362  
 Db 3010 TATTCAATAC-ACCTTTTACCACCTCGACACTACAAAACCAAAATATTTACTAAATAAACA 2952



```

GENERAL INFORMATION:
APPLICANT: Cech, Thomas R.
            Lingner, Joachim
            Nakamura, Toru
            Chapman, Karen B.
            Morin, Gregg B.
            Harley, William B.
            Andrews, William H.
TITLE OF INVENTION: Human Telomerase Catalytic Subunit
NUMBER OF SEQUENCES: 727
CORRESPONDENCE ADDRESS:
ADDRESSER: Townsend and Townsend and Crew LLP
SREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/721,456
FILING DATE: 22-Nov-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-Nov-1997
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-Oct-1996
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-Apr-1997
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-Apr-1997
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-May-1997
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-May-1997
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-Aug-1997
APPLICATION NUMBER: US 08/912,951
FILING DATE: 14-Aug-1997
APPLICATION NUMBER: US 08/915,503
FILING DATE: 14-Aug-1997
APPLICATION NUMBER: WO PCT/US97/17618
FILING DATE: 01-Oct-1997
APPLICATION NUMBER: WO PCT/US97/17885
FILING DATE: 01-Oct-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph Ted
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 015389-002610US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 640:
SEQUENCE CHARACTERISTICS:
LENGTH: 3396 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY:
LOCATION: 1..3396
OTHER INFORMATION: /note= "HRT-encoding sequence employing
alternative codon distributions for
yeast (all genes)"
SEQUENCE DESCRIPTION: SEQ ID NO: 640:
US-09-721-456-640
Alignment Scores:

```

Pred. No.:	0.0132	Length:	3396
Score:	108.00	Matches:	82
Percent Similarity:	38.28%	Conservative:	78
Best Local Similarity:	19.62%	Mismatches:	186
Query Match:	4.52%	Indels:	72
DB:	5	Gaps:	14

  

US-09-910-186A-10 (1-450) x US-09-721-456-640 (1-3396)	
QY 13 AsnSerLeuLeuLysAspPheLeuAsnGluThrPheAsnAsnLeuAspSer-----	30
DB 3048 AACACAGCATGAATCTATAAGCTTGCAACAAACAAATTTTATAAATATTAGTACAAAC	2989
QY 31 -----LysIleLeuSerLeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyr	48
DB 2988 AGTTTGCAAAGAAATTAACTTGCAATCCAAACAAACAAAGAAATGACA-----	2941
QY 49 AsnAlaGluValSerGluGluAspValGlnLeuAsnProIlePhePheProPheAspPhe	68
DB 2940 CAATCTCAAAACACCAACAATTTCTTCTCATATTTTACCAGCTTTAAAAACCTCTATT	2881
QY 69 LysLeuGlySerSerGlyGluAspArgGlyLysVal-----IleValThrGlnAsnGlu	86
DB 2880 AAAAGTCAAGAGCTCTAATAGAGTTCTAGCATAGAGAAATTAATCAGATTGAAAC---	2824
QY 87 AsnIleValTyrAsnSerMetTyrGluSerPheSerIle-----SerPheTyr	102
DB 2823 -----TTCCAAAGTTCTAGTATCCAAACAAACAAACACACCATGG	2785
QY 103 IleArgIleAsnLysTrpValSerAsnLeuProGlyTyrThr-----IleIleAspSer	120
DB 2784 AAACAAACCATGAGCTGGCATTTGACAAACAAACAGTACACCCAAAGCTTCTCTCAAC	2725
QY 121 ValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeuValPheThrLeu	140
DB 2724 TGGAAATTAACACACAGTTTCT-----CAATTAACACACACACCATATTCTGGAAC	2671
QY 141 LysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAsnAla	160
DB 2670 ACCTCTAACCAAGTTCTCAAAAGTTTTCAGCATGAGTCAAAATGTGGAGTAACCAACAA	2611
QY 161 ProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLys	180
DB 2610 AAATCATCAACCAATCTCAACACAAACCATCTCTTAATACCAGCAACAAATTTATT	2551
QY 181 IleTyrIle-----AsnGlyLysLeuIleAsp	189
DB 2550 TTCCATATCACCATACACAAAGACACACAAAGTAGACAAATAGAACCTTGTTGGAAT	2491
QY 190 ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle	209
DB 2490 ACCTTGACATTGAACATAAGATTACCTCTAATTTCAACAGCATGATGACACATAAATCT	2431
QY 210 AsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIle	229
DB 2430 CAAAAA-----AAGTCAACAAACAAACACCA	2407
QY 230 ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn	249
DB 2406 AGAAGCTTCAATCAAGAGAGAGATTGTTCAATACACACAGCATCTCTCAATGGAGAGAT	2347
QY 250 SerLeuGlnTyrThrAsnValValLysAspTyrTrp-----GlyAsnAsp	264
DB 2346 TTCTTGCAATGAGCAACAAATTTGCTCATATATGTTGCAAAATCAGTCAAAAGTAGAAAC	2287
QY 265 LeuArgTyrAsnLysGluTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyr	284
DB 2286 ATGAGATTTTAAAGCTTTTCAACATGACCATGAGCAGCTTTTGAACAAACAGCATATCT	2227
QY 285 AlaAsnSerArgGlnIleVal----PheAsnThrArgAsnAsnAsnAspPheAsnGlu	303
DB 2226 TCTAACACATAAGTATTTTGTGTTTAAATAGAGCAATAACTTCAGTCAATCTATC	2167

QY	96	SerPheSerIleSerPheThrPheIleArgIleAsnLysTrpValSerAsnLeuProGlyTyr	115
Db	2281	AGTATGAA-----ATAGAGATTACAGACATCAAGAATTTGGTTAAGGCGAAA	2328
QY	116	ThrIleIleAsp-----SerValLysAsnAsnSerGlyTrpSerIle	129
Db	2329	AAATTATTAAATTTAGACGAATTAGCACAATCTATGAAAAAACAAC-----	2376
QY	130	GlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsnGluAspSerGluGlnSer	149
Db	2377	-----TTGAAACTAGATGAACGAATTCACAATA2406	
QY	150	IleAsnPheSerTyrAspIleSerAsnAsnAlaProGlyTyrAsnLysTrpPhePheVal	169
Db	2407	GCTACGATTAGCTATAAATAGATAATACTCAACAATTTAATAAA-----	2454
QY	170	ThrValThrAsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAsp	189
Db	2455	-----ATAAAAAAT-----ATAATTCAGATTCGATGAT2484	
QY	190	ThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle	209
Db	2485	GAATTAAAAAGTGCAA-----	2499
QY	210	AsnLysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIle	229
Db	2499	-----	2499
QY	230	ArgAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsn	249
Db	2500	AAAAAATTTTCTGCTATCAGAAAATATTAGCTGTTAA-----	2538
QY	250	SerLeuGluTyrThrAsnValLysAspTyrTrpGlyAsnAspLeuAlgTyrAsnLys	269
Db	2539	-----AATGATTTATAAATTAACACTA2559	
QY	270	GluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGln	289
Db	2560	GAATAATTGTTTAA--ATTATTACTTGGAAAAATAATTTTCAAGCGAATCAACGA2616	
QY	290	IleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyrLysIleIleIle	309
Db	2617	CTTCAATTTATGAGT-----GGAATAGATATCAACTG-----	2649
QY	310	LysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPhe	329
Db	2650	-----ATTAGAGAGAGACCATTTCCTTAGGGTTAAGTGGATTGGAAATTTGACGTGTTT2703	
QY	330	AspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGluThrMetTyrAla	349
Db	2704	GAT-----TTTCATCAATAAATCGAGACACATA2733	
QY	350	AspAsnHisSerThrGluAspIleTyr-----AlaIleGlyLeuArg363	
Db	2734	AGTTCATTTCAGAGGAGAGAACATTTCAAGCATCATCTAGCTTTAGGATTAAGT2793	
QY	364	Glu364	
Db	2794	GAA2796	
RESULT 14			
US-10-240-485-107/c			
; Sequence 107, Application US/10240485			
; GENERAL INFORMATION:			
; APPLICANT: OLEK, Alexander			
; APPLICANT: PIEPENBROCK, Christian			
; APPLICANT: BERLIN, Kurt			
; TITLE OF INVENTION: Diagnosis of Diseases Associated with			
; TITLE OF INVENTION: Metastasis			
; FILE REFERENCE: 5013.1007			
; CURRENT APPLICATION NUMBER: US/10/240.485			
; CURRENT FILING DATE: 2002-10-02			
; PRIOR APPLICATION NUMBER: PCT/EP01/03970			



Db 6291 CACTACATTATTACGCTCAAC-----ATAACACTCAATAT 6256  
QY 173 AsnAsnMetMetGlyAsnMetLysLeu-TyrIleAsnGlyLeuIleAspThrIle 192  
||||| : : : : :  
Db 6255 AACACATTAATACTATCAAAAAAATATATACCAAGCGAAGTATACAACTACTAA 6196  
QY 192 sValLysGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIle----- 209  
| : : : :  
Db 6195 AAAAAACAACCTCCATACATTAATCAAAAAATTAATTAACCCATCGAATCCACTAAAA 6136  
| : : : :  
QY 209 ----- 209  
Db 6135 AACTATCCATACACAACAACATCCACAACCAAAAAACAAAAACAAATTAACCTT 6076  
QY 210 -----AsnLysIleProAspThrGlyLeuIle----- 218  
| : : : :  
Db 6075 ACACACAACCTTATTATACAAAAAATACATCCCTTTACTCCCTCTTATTATAATC 6016  
| : : : :  
QY 219 -----ThrSerAspSerAspAsnIleAsnMetTrpIle 230  
| : : : :  
Db 6015 CAAACCTTAACCTCTGCTACTCCACAACCTCTCAAAATCACCTTAACCTC----- 5964  
QY 230 gAspPheTyrIlePheAlaLysGluLeuAspGlyLysAspIleAsn---IleLeuPheAs 249  
| : : : :  
Db 5963 -AACTTCCCGTTCACAATTTCGAAACCAACCTCCCTTAATATTAACTACATATTATTCAC 5905  
| : : : :  
QY 249 n-----Se 250  
Db 5904 ATTACTTCATATTATTTACCTCAATATATTTAAAAACGCACACTAAACCATATTTTA 5845  
QY 250 rIeuGlnTyrThrAsnValValLysAspTyrTrpGlyAsnAspLeu----- 265  
| : : : :  
Db 5844 TTTAAATTTCAATATTCCTCCAAAATTTCTCCCTTTTAAACTACTATATTAAAT 5785  
| : : : :  
QY 266 -ArgTyrAsnLysGluTyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyr-A 285  
| : : : :  
Db 5784 CCAATATCTTCCCATTTAC-----TTAACAATAACATTTTAAAT-----CACCC 5740  
QY 285 laAsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsnGluGlyT 305  
| : : : :  
Db 5739 CTAAACACTACTCAAAAAATACACACCAACCAATTAATCATCAATATATAAAATCCCTACCT 5680  
| : : : :  
QY 305 rLys---IleIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyG 324  
| : : : :  
Db 5679 ATAAACCTATTCTATCCCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5627  
QY 324 lYAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeuPheMetLysA 344  
| : : : :  
Db 5626 -----CTCCTTTTATTCTCA---AATCTCATATATCTATTAAATCAATATAACCTAAAA 5575  
QY 344 sGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeuArgG 364  
| : : : :  
Db 5574 ATAAACCCGTC-----TCAAACGCTTCCCAATTTAATTTATATCTCTATTACATAAAC- 5522  
QY 364 luGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPromet----- 380  
| : : : :  
Db 5521 -----TCCACAAATATCTCATACAAATATACAAATATATATATATATATATAT 5479  
QY 381 -----AsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnG 397  
| : : : :  
Db 5478 AACATTACCGAATAAT-----TATTCTCCAAACAAATAACTAAAAACAATCTAAATA 5425  
QY 397 lYGluAsn 399  
| : : : :  
Db 5424 TTCATAAT 5417

Search completed: November 7, 2002, 20:05:40  
Job time : 62 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:41:56 ; Search time 43 Seconds

(without alignments)  
1006.058 Million cell updates/sec

Title: US-09-910-186a-10

Perfect score: 2391

Sequence: 1 MTIPNFSTYNNLLKDI.....NYASLLESTHWGFVPVSE 450

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2386	99.8	1291	2 S46431	botulinum neurotox
2	2386	99.8	1291	2 A49777	botulinum neurotox
3	1823	75.2	1285	2 S70582	botulinum neurotox
4	867.5	36.3	1276	2 S11435	botulinum neurotox
5	604.5	25.3	1297	2 S39791	neurotoxin - Clost
6	594.5	24.9	1252	2 S21178	botulinum neurotox
7	594	24.8	1251	2 JH0256	botulinum neurotox
8	593	24.8	1291	2 I40631	non-proteolytic bo
9	586.5	24.5	1296	1 BTCLAB	bontoxilysin (BC 3
10	581	24.3	1291	1 A48940	bontoxilysin (BC 3
11	569	23.8	1268	2 S33411	botulinum neurotox
12	554.5	23.2	1296	2 I40645	botulinum neurotox
13	549.5	23.0	1274	2 I40813	neurotoxin type F
14	425	17.8	1315	1 BTCLTN	tentoxilysin (BC 3
15	281.5	11.8	367	2 S48106	neurotoxin type E
16	279	11.7	366	2 S48110	neurotoxin type F
17	245.5	10.3	369	2 S48109	neurotoxin type F
18	230.5	9.6	1162	2 A47708	progenitor toxin n
19	230.5	9.6	1162	2 I40817	botulinum toxin no
20	213	8.9	1193	2 S68218	botulinum neurotox
21	199	8.3	1165	2 I40644	botulinum neurotox
22	199	8.3	1196	2 JQ1467	toxin, nontoxic co
23	199	8.3	1196	2 S46430	botulinum neurotox
24	191.5	8.0	1193	2 JQ4901	toxic-nonhemagg
25	174.5	7.3	960	2 S72284	DNA-directed RNA p
26	171.5	7.2	1844	2 D71612	hypothetical prote
27	170.5	7.1	398	1 R3BYM1	ribosomal protein
28	168.5	7.0	1817	2 H71611	probable secreted
29	168	7.0	4688	2 F62885	hypothetical prote

ALIGNMENTS

RESULT 1

S46431

botulinum neurotoxin C1 - Clostridium botulinum phase 1C (strain C 468)

N.Alternate names: BoNT/C1 protein

C.Species: Clostridium botulinum phase 1C

A.Variety: strain C 468

C.Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 16-Jul-1999

R.Accession: S46431; S49107

R.Hauser, D.; Eklund, M.W.; Boquet, P.; Popoff, M.R.

Mol. Gen. Genet. 243, 631-640, 1994

A.Title: Organization of the botulinum neurotoxin C1 gene and its associated non-toxi

A.Reference number: S46426; MUID:94301293; PMID:8028579

A.Accession: S46431

A.Status: nucleic acid sequence not shown; translation not shown

A.Molecule type: DNA

A.Residues: 1-1291 <HAU>

A.Cross-references: EMBL:X72793; NID:g516171; PIDN:CAA51313.1; PID:g516175

A.Experimental source: strain C 468

A.Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1993

C.Superfamily: tetanus toxin

Query Match 99.8%; Score 2386; DB 2; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 1.7e-129;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	TIPNFSTYNNLLKDIINEYFNNDISKILSCQNRKNTLVDTSGYNAEVSERGDYQLN	61
Db	843	TIPNFSTYNNLLKDIINEYFNNDISKILSCQNRKNTLVDTSGYNAEVSERGDYQLN	902
QY	62	PIPFDFKLGSSGEDRGKVIYVTQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV	121
Db	903	PIPFDFKLGSSGEDRGKVIYVTQENIYVNSMYESFSISFWIRINKWVSNLPGYTIIDSV	962
QY	122	KNSGWSIGIISNLFVFLKQNEDESEQINSYDISNAPGYNKWFVVTNNMGNKKI	181
Db	963	KNSGWSIGIISNLFVFLKQNEDESEQINSYDISNAPGYNKWFVVTNNMGNKKI	1022
QY	182	YINGKLIDITIKVELTGINFSTITFEINKIPDGLITSDSDNINMIRDFYIFAKELDG	241
Db	1023	YINGKLIDITIKVELTGINFSTITFEINKIPDGLITSDSDNINMIRDFYIFAKELDG	1082
QY	242	KDINILFNSLQTVVVKDYGWNLDRYKNEYMVNIDYLNRYMYANSRQIVENTRRNND	301
Db	1083	KDINILFNSLQTVVVKDYGWNLDRYKNEYMVNIDYLNRYMYANSRQIVENTRRNND	1142
QY	302	NEGYKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMADNHNSTEDIYAIG	361
Db	1143	NEGYKIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNETMADNHNSTEDIYAIG	1202
QY	362	LRQTKDINDNIIFOIQPNNTYYVQASQIFKSNFNGENISGICSGTYFRFLGGDWTRHN	421
Db	1203	LRQTKDINDNIIFOIQPNNTYYVQASQIFKSNFNGENISGICSGTYFRFLGGDWTRHN	1262

hypothetical prote  
uncharacterized pr  
probable periplasm  
hypothetical prote  
hypothetical prote  
hypothetical prote  
probable membrane  
ORF MSV152 probabl  
hypothetical prote  
hypothetical prote  
ORF MSV156 hypothe  
hypothetical prote  
hypothetical prote  
serine/threonine-s  
hypothetical prote



QY 302 NEGYKIIIKIRGNTNDRVRGGDIIYFDMTINKKAYNFMKNETMYADNHSTEDYIAG 351  
 Db 1143 NEGYKIIIKIRGNTNDRVRGGDIIYFDMTINKKAYNFMKNETMYADNHSTEDYIAG 1202  
 QY 362 LRQTKDINDNIIFQIQPMNNTYYIASQIFKSNFNGENISGICSIGTYRFLGDDWYRHN 421  
 Db 1203 LRQTKDINDNIIFQIQPMNNTYYIASQIFKSNFNGENISGICSIGTYRFLGDDWYRHN 1262  
 QY 422 YLVPTVKQGNYSALLSTSTHGWGFVPVSE 450  
 Db 1263 YLVPTVKQGNYSALLSTSTHGWGFVPVSE 1291  
 RESULT 3  
 S70582  
 botulinum neurotoxin type Dsa precursor - Clostridium botulinum phage d-sa  
 C:Species: Clostridium botulinum phage d-sa  
 A:Note: host Clostridium botulinum type D (strain South Africa)  
 C:date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 20-Jun-2000  
 C:Accession: S70582  
 R:Morilishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.  
 Biochim. Biophys. Acta 1307, 123-126, 1996  
 A:title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C  
 A:reference number: S70582; MUID:96283801; PMID:8679691  
 A:accession: S70582  
 A:status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1285 <MOR>  
 A:Cross-references: EMBL:D38442; NID:gl1374775; PIDN:BAA07477.1; PID:gl1374776  
 C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit  
 a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic me  
 C:Superfamily: tetanus toxin  
 C:Keywords: disulfide bond; neurotoxin; transmembrane protein  
 F:1-447/Product: botulinum neurotoxin type Dsa light chain #status predicted <MAT1>  
 F:448-1285/Product: botulinum neurotoxin type Dsa heavy chain #status predicted <MAT1>  
 Query Match 76.2%; Score 1823; DB 2; Length 1285;  
 Best Local Similarity 77.7%; Pred. No. 3.8e-97;  
 Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;  
 QY 2 TIFNIFSYNNLSLLKDIIINEYFNINDSKILSLQNRKNTLVDTSGYNAEVSEGDVQLN 61  
 Db 839 TIFNIFSYNNLSLLKDMINEYFNINDSKILSLQNRKNTLVDTSGYNAEVSEGDVQLN 898  
 QY 62 PIFPFDKLGSSGDRKVIIVTONENIVYNSWESFISFWIRINKWVSNLPGYTIIDSV 121  
 Db 899 PIFPFDKLGSSGDRKVIIVTONENIVYNSWESFISFWIRINKWVSNLPGYTIIDSV 958  
 QY 122 KNSGWSIGIISNPLVFLTKQNSEQDINFSYDINSNAPYKNKFFVTVTNNMGNMKI 181  
 Db 959 KNSGWSIGIISNPLVFLTKQNSEQDINFSYDINSNAPYKNKFFVTVTNNMGNMKI 1018  
 QY 182 YINGKLIDTIVKVELTGINSKITPTPIINKIPDGLTSDSDINMWRDYPYFAKELDG 241  
 Db 1019 YINGKLIDTIVKVELTGINSKITPTPIINKIPDGLTSDSDINMWRDYPYFAKELDD 1078  
 QY 242 KDNILFNSLQYTNVVDYGNLDRLRYKEYVMYNDILNRYMTANSRQIVFNVRNNNDF 301  
 Db 1079 KDNILFNSLQYTNVVDYGNLDRLRYKEYVMYNDILNRYMTANSRQIVFNVRNNNDF 1138  
 QY 302 NEGYKIIIKIRGNTNDRVRGGDIIYFDMTINKKAYNFMKNETMYADNHST-EDIYAI 360  
 Db 1139 NEGYKIIIKIRGNTNDRVRGENVLVENTIDNKQYSL-----GMYPKSRNLTGDLVPL 1193  
 QY 361 GLRQTKD-INDNIIFOIQPMNNTYYIASQIFKSNFNGENISGICSIGTYRFLGDD-WY 418  
 Db 1194 GALDQPMDETRKYGSLFIQPCNFDYASQLFLSSNATTNRLGLTSLIGSYFKLGGDYWF 1253  
 QY 419 RHNLYPTVKQGNYSALLSTSTHGWGFVPVSE 450  
 Db 1254 NHEYLIVPIKIEHYASLLESTSTHWFVPASE 1285

```
RESULT 4
S11455
botulinum neurotoxin type D - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 18-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C:Accession: S11455
R:Binz, T.; Kurazono, H.; Popoff, M.R.; Eklund, M.W.; Sakaguchi, G.; Kozaki, S.; Kriegl
Nucleic Acids Res. 18, 5556, 1990
A:Title: Nucleotide sequence of the gene encoding Clostridium botulinum neurotoxin type
A:Reference number: S11455; MUID:91016893; PMID:2216736
A:Accession: S11455
A>Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 1-1276 <BIN>
A:Cross-references: EMBL:X54254; NID:g40395; PIDN:CAA38175.1; PID:g40396
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin

Query Match 36.3%; Score 867.5; DB 2; Length 1276;
Best Local Similarity 43.2%; Pred. No. 3e-42;
Matches 195; Conservative 83; Mismatches 146; Indels 27; Gaps 11;

Qy 2 TIPNIFSYTNSLLKDIINEYFNNDKSLQKLNKNTLVDTSGYNAEVSEGGVQLN 61
Db 839 TWPFNIFSYTNSLLKDIINEYFNNDKSLQKLNKNTLVDTSGYNAEVSEGGVQLN 898
Qy 62 PIFPFDFKLGSGEDRGKVIYVTONENIVYNSMYESFISFWIRNKWVSNLPG-YTIIDS 120
Db 899 TYTNDKFLSSGD---KIIYVNLNNILYSAIVENSVSFMIKSKDLTNSHNEYIINS 955
Qy 121 VKNNSGWSIGIISNLFVTLTKQNEDESQSINFSYDVISNAPGY-NKWFVTVTNMAGNM 179
Db 956 LEQNSGWSIGIISNLFVTLTKQNEDESQSINFSYDVISNAPGY-NKWFVTVTNMAGNM 1015
Qy 180 KIYINGKLIDIKVKEITGTFNFKITPDGTGLITSDSDNINMIRDPYIPAKEL 239
Db 1016 KLYINGELKQSKIEDLEVKLDKTIYFVIDENTID-----ENQMLARDENIFSKEL 1067
Qy 240 DGKIDINILFNSLQTVVYKQWGNLDYKRYKMYWNIYDLYNRYMYANSQIVFNTRRNN 299
Db 1068 SNEDINIVYEGQILRNVIKDYGNPLKEDFYIINDYIDRYIAPESNVLVLVQVPDRS 1127
Qy 300 DFNEGYKIIIRIRGNTDNRVRGSDILYFDMTINNKAYNLFKMETVADNH---STED 356
Db 1128 KLYTGNPTITKSDKPNYSRLNGDNIILHLYNSRKYMIIRDYTIYATOGGECQNC 1187
Qy 357 IYAGLREQTQKIDINDNIFQIQPNNTYYVYASQIFKSNFNGENISGICSI-GTYRFLGG 415
Db 1188 VYALKLQSNLGNYGIG-IFSIKNIYVSNKYCSQIF-SSFR-ENTMLLADIYKWPRES--- 1241
Qy 416 DWYRHNLYVPTVKQGNVYASLLESTSTWGFV 446
Db 1242 --FKNATPVAVT--NYETKLLSTSSFWKFI 1268

RESULT 5
S39791
neurotoxin - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C:Accession: S39791
R:Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A:Title: Nucleotide sequence of the gene coding for Clostridium botulinum
A:Reference number: S39791; MUID:94052745; PMID:8268233
A:Accession: S39791
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1297 <CAM>
A:Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C:Superfamily: tetanus toxin
C:Keywords: neurotoxin
```

```
Query Match 25.3%; Score 604.5; DB 2; Length 1297;
Best Local Similarity 32.0%; Pred. No. 4e-27;
Matches 156; Conservative 87; Mismatches 172; Indels 73; Gaps 16;

Qy 2 TIPNIFSYTNSLLKDIINEYFNNDKSLQKLNKNTLVDTSGYNAEVSEGGVQLN 61
Db 840 SIPDLSLYTKDTLILQVFNYYISNISSNAILSLSYRGGRLLDSGYGATMNVGSDVFN 899
Qy 62 PIFPFDFKLGSGEDRGKVIYVTONENIVYNSMYESFISFWIRNKWVSN-----LPQY 115
Db 900 DIGNQPKLANS--ENSNTIAHQSKFVYVYDSMFEDSINFVWETPKYNNNDIQTLYQNEY 957
Qy 116 TIIDSVKNNSGWSIGIISNLFVTLTKQNEDESQSINFSYDVISNAPGY-NKWFVTVTN 174
Db 958 TIISCIKNNSGWSIGIISNLFVTLTKQNEDESQSINFSYDVISNAPGY-NKWFVTVTN 1017
Qy 175 MGNMKNYINGKLIDITKVKELTGINFSEKITITFEINKIPDTGLITSDSDNINMIRDPYI 234
Db 1018 RLGNTANYIINGSLKSKILNLDRISSNDIDFKLNCIDTDTKF-----VWIKDFNI 1069
Qy 235 FAKELDGKIDINILFNSLQTVVYKQWGNLDYKRYKMYWNIYDLYNRYMYANSRQIVFT 294
Db 1070 FGRELNATEVSSLYWIOSNTLKDQWGNPLRDTQYLFNQCMQNIYKYFSKASMGET 1129
Qy 295 RRNNNDPNE-----GYKIIIKRI---RGNTNDTRVGGDILYFDMTINNKAYNLF 341
Db 1130 APRTN--FNNAINYNQNLGLRFLFIKKASNRNINNDNIVREGDIYLN----- 1178
Qy 342 MKNETMADNHSDEIYVAGLREQTQKIDINDNIFQIQPNNTYYVYASQI---FKSNF 395
Db 1179 -----DNISDES-YRVTVLVNSKEIQQLF--LAPINDPTFYDVLQIKYKYKTYIY 1227
Qy 396 NGENIS-----GICISIGTYRFLRGSDYRHH-NYLVTPTVKQGNVYASLLESTS-----TH 442
Db 1228 NCQILCEKDKTKFTGLFGIGKFKVDYGVVMDYDNYF--CISQWYLRRISENINKRLGNC 1285
Qy 443 WGFVPVSE 450
Db 1286 WQFIPVDE 1293

RESULT 6
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C:Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A:Title: The complete amino acid sequence of the Clostridium botulinum type-E neuroto
A:Reference number: S21178; MUID:92174922; PMID:1541280
A:Accession: S21178
A:Molecule type: DNA
A:Residues: 1-1252 <WHE>
A:Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R:Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific
A:Reference number: S48103; MUID:94013372; PMID:8408542
A:Accession: S48107
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 616-982 <CAM>
A:Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A:Title: Sequences of the botulinum neurotoxin E derived from Clostridium botulinum t
A:Reference number: JH0257; MUID:92181428; PMID:1543481
A:Accession: JH0257
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
```

RESULT 7  
JH0256  
botulinum neurotoxin type E precursor - Clostridium butyricum  
C:Species: Clostridium butyricum  
C:Date: 30-Jun-1992 sequence\_revision 15-May-1998 #text\_change 16-Jul-1999  
C:Accession: JH0256; S16145  
R:Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.

```

RESULT 8
140631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C:Species: Clostridium botulinum
C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C:Accession: I40631; S48103; S48104; S36015
R:Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curt. Microbiol. 28, 101-110, 1994
A:Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botu

```



C/Genetics: A:Residues: 634-994 <CAM>  
C/Gene: act; bota  
C/Function: A:Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1.; PID:g407783  
C:Superfamily: tetanus toxin  
C:Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associated  
F:2-444/Product: botoxilysin A light chain #status experimental <LIGHT>  
F:445-1236/Product: botoxilysin A heavy chain #status experimental <LIGHT>  
F:223-227/Binding site: zinc (His) #status predicted  
F:224/Active site: Glu #status predicted

Query Match 24.5%; Score 586.5; DB 1; Length 1296;  
Best Local Similarity 31.6%; Pred. No. 4.4e-26;  
Matches 149; Conservative 82; Mismatches 186; Indels 55; Gaps 11;

QY 3 IPNLFYSYNNLLKDIINEYFNNDKSKILSLQNRKNTLVDFSGYNVSEGGVQLNP 62  
DB 849 IPFQSKYVDNQRLLSTTFEIKNIWTSILNLYRSNHLIDLSRYASKINGSKVNP 908  
QY 63 IPFPD---FKLGSGEDRGKVVIVQTNENIVYNSMYESFSFIRINKVSNLP---GYT 116  
DB 909 IDKNQQLFNLESS-----KTEVLKNAIVYNSMYENFSEFIRIPKVFNSISLNNEYT 963  
QY 117 IDSVKNNSGWIGIISNLFVTLKONEDSEOSNFESYD:SNNAPOY-NKRFVFTVNNM 175  
DB 964 IINCENNSGWKYSUNYEITWLTQDTQEKQVRVYKYSQMINISYINRWLFVITNRR 1023  
QY 176 MGNMKYIYNGKLIDPIKYEKLTGINSKTIFFTEINKIPDPGLTISDSNINWIRDFYIF 235  
DB 1024 LNSKIYINGRIDOKPISNLGNTHASNNIMFKLDGCRDP-----HRYTWIKYFNLF 1075  
QY 236 AKELGKDNILFNSLOTVNVKDYWDGNDLRYNKYYWIDLYNRYMVS-----287  
DB 1076 DRELNEKEIKLDYDQNSGILKDFWDYQYDYPYMLNLYDPKYYVDVNNVGVIGYMY 1135  
QY 288 ---ROIVNTR-RNNDFNENGYKIIKIRINGNTNDFPRVGGDILYFDMTINKNAYNLFM 342  
DB 1136 LKPGRSVMTTYIYNSLLYAGTRFFIKKYSAGKNDIVNRNDRVYINVVVKNKEYL-- 1193  
QY 343 KNETYADNHSFEDYATGLREQTKDNDNIIFQIOPMNTYTYASQIFKSNFNGENTSG 402  
DB 1194 -----ATNASQAGVEKILSALEPVDGNLSQVVMKSKNDQGITNKC-KMLQDNNGND 1246  
QY 403 ICIGTYRFLGGDWYRNLYVPVTKGCVNYSALLESTS-----THWGFVPSVE 450  
DB 1247 IGFIFGHQP-----NNIAKLVASWYNRQIERSRSLGCSWFEFIPVDD 1289

RESULT 10  
A48940  
botoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum  
C/Alternate names: botulinum neurotoxin type B (BoNT/B)  
C/Species: Clostridium botulinum  
C/Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
C/Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08574  
R:Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, N.P.  
Appl. Environ. Microbiol. 58, 2345-2354, 1992  
A:Title: Molecular cloning of the Clostridium botulinum structural gene encoding the type  
A:Reference number: A48940; MUID:923384550; PMID:1314783  
A:Accession: A48940  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1291 <WRE>  
A:Cross-references: GB:M81186; NID:g144734; PIDN:AAA3211.1; PID:g144735  
A:Experimental source: type B, Danish  
A:Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica  
R:Campbell, K.D.; Collins, M.D.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993  
A:Title: Gene probes for identification of the botulinum neurotoxin gene and specific id  
A:Reference number: S48103; MUID:94013372; PMID:8408542  
A:Accession: S48105  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 634-994 <CAM>  
A:Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1.; PID:g407783  
A:Experimental source: proteolytic type B, strain NCTC 7273  
R:Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.  
submitted to the EMBL Data Library, April 1992  
A:Description: Partial amino acid sequence of botulinum neurotoxin type B and compari  
F:2-444/Product: botoxilysin B light chain #status experimental <LIGHT>  
F:445-1236/Product: botoxilysin B heavy chain #status experimental <LIGHT>  
F:230-234/Binding site: zinc (His) #status predicted  
F:231/Active site: Glu #status predicted

Query Match 24.3%; Score 581; DB 1; Length 1291;  
Best Local Similarity 32.9%; Pred. No. 9e-26;  
Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;

QY 3 IPNFTSTNNSLLKDIINEYFNNDKSKILSLQNRKNTLVDTSGYNVSEGGVQLNP 62  
DB 836 MPFDLSIYNTDILTFMFKYNSSEILNLIYRKNLIDLSYGAKEVYDGVGLND 895  
QY 63 IPFPDFKLGSGEDRGKVVIVQTNENIVYNSMYESFSFIRINKV---VSNL--PGYT 116

Db 896 ---KNQFKLTSSA---NSKLRVTONQNIENFSLDFSVFWIRIPKYANDGIQYIHNEYT 951  
QY 117 IIDSVKNSGWSIGIISNPLVFTLKQNEDESQSINFSYDISNNAPGY-NKWFVYVYNNM 175  
Db 952 IINCNNKNSGWSIGIRNRRIIWTLDINGKTKSVFFEYNIREDISEINRWFFVTINN- 1010  
QY 176 MGNMKIYINGKLIDIPKVKELTGTFNFKSTITFEINKIPDPTGLTSDSDNIN-MWIRDFYI 234  
Db 1011 LNNAKIYINGKLESNTDKIDREVIANGEIIFKLD-----GDIDRQPIWAKYFSI 1061  
QY 235 FAKELDGDNDINFLNSLOYTNVVDYWGNDLRKYNEYMYMVIDLNRMYMANS-----RQI 290  
Db 1062 FNTLSQSNIEERYKIQSYSEYKLDWFCNGLMYNKEYMFMAGNKNSYIKLKXKDSVGEI 1121  
QY 291 VFNTRNNN-----DFNEGYKIIKIRGNT--NDTRVRGGDILYFD-MTINKN--- 336  
Db 1122 LFRKYNQNSKYINVRDLYIGEKFIIR-KNSQSINDDIVRKEDYIYLDFFNLNWEVRV 1180  
QY 337 -AYNLFMKNETMADNHSTEDIYAIGLEQTKDINDNIIFOIQPMNNYYIASQ-IFKSN 394  
Db 1181 YVYKFKKEE-----EKLFLAPISDSDEFYN--TQIKKEYDEQPYSCQLLFKK- 1227  
QY 395 FNGENISGICSGTVRFRLLG-----DWYRHNYLVPTVKOGNYASLLESTST 441  
Db 1228 -DEESTDELGLIGHRFYSGIVPEYKDYFCISKWY-----LAEVRKPYNLKL---GC 1278  
QY 442 HWGFVPVSE 450  
Db 1279 NMQFIPKDE 1287

RESULT 11  
S33411  
botulinum neurotoxin type F - Clostridium baratii  
C:Species: Clostridium baratii  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Jul-1999  
R:Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.; FEMS Microbiol. Lett. 108, 175-182, 1993  
A:Title: Nucleotide sequence of the gene coding for Clostridium baratii type F neurotoxin  
A:Reference number: S33411; MUID:93252228; PMID:8486245  
A:Accession: S33411  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1268 <THO>  
A:Cross-references: EMBL:X68262; NID:949138; PIDN:CAA48329.1; PID:g49139  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 23.8%; Score 569; DB 2; Length 1268;  
Best Local Similarity 32.7%; Pred. No. 4.3e-25;  
Matches 154; Conservative 82; Mismatches 167; Indels 68; Gaps 17;

QY 2 TPFNIFSTNNSLKLDINEYFNNDISKILSQNRKNTLVDTSGYNAEYSEGDVQLN 61  
Db 832 SIPFELSEYTNPKDILHILRFYKRIIDSSILNKMKNYENRFDISGYSNISINGDIYI 891  
QY 62 PIPFPFKLGGSGEDRGKVIYTONENIVYNSMYESFSISFWIRINKW--VSNLPG-VYII 118  
Db 892 STNRNCPGYISS--RLSEWNTQNTNIYNSRYQNFVSFWVRIPKYNLKNLNNETII 949  
QY 119 DSVK-NNSGWSIGITSNPLVFTLKQNEDESQSINFSYDISNNAPGYNKWFVYVYNNM 173  
Db 950 NCMRRNNSGWSIGISLNNIITWLTQDNTGNOKLVFN--TQIDISDY---INKWTFVITN 1006  
QY 174 NMGNMKIYINGKLIDIPKVKELTGTFNFKSTITFEINKIPDPTGLTSDSDNINMWIRDFY 233  
Db 1007 NRUGHSKLTINGLNDQKSLNGLNHWDDNLKFKVGCNDTRYV-----GIRYFK 1057  
QY 234 IFAKELDGDNDINFLNSLOYTNVVDYWGNDLRKYNEYMYMVIDLNRMYMANSRQVFN 293  
Db 1058 IFNMELDKTEIETLYHSEPDSTILKDFWGNVLYNKKYIYLLNLLKPNMSVTKNSDILNIN 1117

QY 294 TR-----NNDFNEGYKIIKIRGNTN---DTRVRGGDILYFDMTINKKAYNL 340  
Db 1118 RORGYSKTNIFSNARLYTGVVEIIRKVGSTDTSTNDFNVRKNDTVIINVDGNEIQL 1176  
QY 341 FMKNETMADNHSTEDIYAIGLE-OTKNDINDNIIFOIQPMNNYYIASQIFKSNFNGEN 399  
Db 1177 -----YADYSTSAVEKTIKLRISNSYNSN--QMIWDSIGDNDCTMNFXTN-NGND 1225  
QY 400 ISGICSGITVFRLL-----GGDWYRHNYLVPTVKOGNYASLLESTSTHWGFV 446  
Db 1226 -----IGLGFHLNVLVYASWYKYNIRNTRNNGCF-----WSFI 1260

RESULT 12  
I40645  
botulinum neurotoxin type A - Clostridium botulinum  
C:Species: Clostridium botulinum  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 16-Jul-1999  
C:Accession: I40645  
R:Williams, A.; East, A.K.; Lawson, P.A.; Collins, M.D.; Res. Microbiol. 144, 547-556, 1993  
A:Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A  
A:Reference number: I40645; MUID:94143603; PMID:8310180  
A:Accession: I40645  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1296 <RES>  
A:Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 23.28%; Score 554.5; DB 2; Length 1296;  
Best Local Similarity 30.8%; Pred. No. 3e-24;  
Matches 147; Conservative 81; Mismatches 183; Indels 67; Gaps 12;

QY 3 IPFNIFSTNNSLKLDINEYFNNDISKILSQNRKNTLVDTSGYNAEYSEGDVQLNP 62  
Db 849 IPFQLSKYVDNKKLLSTETEKIVNTSILSIYKKDLDLSRYGAKINIGRVDYDS 908  
QY 63 IPFPFKLGGSGEDRGKVIYTONENIVYNSMYESFSISFWIRINKWVSNL---PGYTIID 119  
Db 909 IDKNQIKLINLESSTIEVIL--KNAIVNSMYENFSTFWIKIPKFSKINLNNETIIN 966  
QY 120 SVKNSGWSIGITSNPLVFTLKQNEDESQSINFSYDISNNAPGY-NKWFVYVYNNMGN 178  
Db 967 CIENSGWKVSLNYGEIITWLTQDNKQNIQVYFKYSQWNLSDYINENIWFVITNNELTK 1026  
QY 179 MKIYINGKLIDIPKVKELTGTFNFKSTITFEINKIPDPTGLTSDSDNINMWIRDFYIFAKE 238  
Db 1027 SKIYINGRLIDQKPISNLGN-HASNKIMFKLDGCR-----PRYIMKYFNLPDKE 1078  
QY 239 LDKDINTILNSLOYTNVVDYWGNDLRKYNEYMYMVIDLNRMYMANS----- 287  
Db 1079 LNEKIKLDYSQSGSILKDFWGNLYQDPYVYMLNLFDPKDYVDNNGIRGYMYLKG 1138  
QY 288 --RQIVFNRNNDFNEGYKIIKIRGNTNDRVRGGDILYFDMTINKKAYNLFMKNE 345  
Db 1139 PRGSVVTNIYLSNLTLEGTFKFLIKKVASGNEEDIVRNDRVYINVYVKNKEYRL----- 1193  
QY 346 TWYADNHSTEDIYAIGLEQTKDINDNIIFOI-----OPMNTYYIASQIFKSNFNG 397  
Db 1194 ----ATNASOAGVEKLTLSALEIPDVGN--LSQVYVWVKSKDDOGIRNK-----CKMNLQD 1241  
QY 398 ENISGICSGITVFR-----RLGGDWYRHNYLVPTVKOGNYASLLESTSTHWGFVPSVSE 450  
Db 1242 NNGNDIGFIFGLHYDIAKLVASNWN-----RQVGRKAS--RTFGCSWEFIPVDD 1289

RESULT 13  
I40813  
botulinum neurotoxin type F - Clostridium botulinum  
C:Species: Clostridium botulinum

C>Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999  
C:Accession: I40813; S48108  
R:East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E  
FEMS Microbiol. Lett. 96, 225-230, 1992  
A:Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.  
A:Reference number: 140644  
A:Accession: I40813  
A:Status: preliminary; translated from GB/EMBL/DDJ  
A:Molecule type: DNA  
A:Residues: 1-1274 <RES>  
A:Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867  
R:Campbell, K.D.; Collins, M.D.; East, A.K.  
J. Clin. Microbiol. 31, 2255-2262, 1993  
A:Title: Gene probes for identification of the botulin neurotoxin gene and specific id  
A:Reference number: S48103; MUID:94013372; PMID:8408542  
A:Accession: S48108  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 634-1002 <CAM>  
A:Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789  
C:Superfamily: tetanus toxin  
C:Keywords: neurotoxin

Query Match 23.08; Score 549.5; DB 2; Length 1274;  
Best Local Similarity 31.68; Pred. No. 5.7e-24;  
Matches 147; Conservative 83; Mismatches 180; Indels 55; Gaps 15;

QY 2 TIFNIFSYNNLSKLDINIFYNNINDSKILSONKNTLVDTSGYNAEVEGVDQLN 61  
DB 840 SIPFELSYNTDKILIFNRYLKIKKIDSSILDMRYENKFKIDISGYSNINGNVYIY 899  
QY 62 PIFPDKLGGSGEDRGKVIYTONENIYVNSMFYSFISFWIRNKW--VSNLPGVTII 118  
DB 900 STNRNQF--GIYNSLSEVIAQNDIIYNSRYQNFISFWIRPKHYKPMHNRXTII 957  
QY 119 DSV-KNNSGWSIG--IISNLFVLTQKNESEISNFSYDIDSNAPGY-KNMFVFTVN 173  
DB 958 NCMGNNNSGKISLRTVRCDEIITLQDTSGNKENLFRVEELNRIYNNKIFVTIN 1017  
QY 174 NMGNKVIYNGKLDTKVRELGTGINSKTIPTKINKIPTDGLTSDSNINMWRDFY 233  
DB 1018 NRIISRYIYNGNLIVKESISNLGDIHVSMDILFK-----IVGCDTETVYGIIRFK 1068  
QY 234 IFAKELGKDINILFNSQYTNVVKDYNGDLRYNKEYIYNNIDILNRYNANSRQIVFN 293  
DB 1069 VFNTLDTKEITLVSNEPDPSILKNYGNVLLYNNKKYLLNLRKDKYITLNSGILNIN 1128  
QY 294 TRRN-----NDFNKGKIIIR---IRGNTDVRVRCGDILYDMDTNNKAYNLFM 342  
DB 1129 QQRGVTEGSVFLNYKLYGVEVIRKNGPIDISNTDFNVRKNDLAYINVVDRGVEYRL-- 1186  
QY 343 KNETYADNHSTEDIVAGLRQTRKDINDNIIFQIPMNNPYYVASQIFKSNFNGENISG 402  
DB 1187 ----YADTKSEKII-----RTSNLNDLS-LGQIIVMDSIGNCTMNFQNN-NGSN-- 1231  
QY 403 ICSIGYRFR---LGGDWYRNLYLPVTKQGNVYASLLESSTHW 443  
DB 1232 ---IGLLGFHNNLVASSWYNNIRNRNTSSNGCFWSSI-SKENGW 1272

RESULT 14  
BTCLTN  
tentoxilysin (EC 3.4.24.68) precursor - Clostridium tetani  
N:Alternate names: tetanus neurotoxin  
C:Species: Clostridium tetani  
C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 03-Jun-2002  
C:Accession: A25689; A25194; B25194; A60759; S69348; S09364  
R:Eisel, U.; Jarausch, W.; Goretzki, K.; Henschen, A.; Engels, J.; Weller, U.; Hudel, M.  
EMBO J. 5, 2495-2502, 1986  
A:Title: Tetanus toxin: primary structure, expression in E. coli, and homology with botu  
A:Reference number: A25689; MUID:87053614; PMID:3536478  
A:Accession: A25689  
A:Molecule type: DNA

A:Residues: 1-1315 <EIS>  
A:Cross-references: GB:X04436; NID:g40769; PIDN:CAA28033.1; PID:g40770  
R:Fairweather, N.F.; Lyness, V.A.  
Nucleic Acids Res. 14, 7805-7812, 1986  
A:Title: The complete nucleotide sequence of tetanus toxin.  
A:Reference number: A25757; MUID:87040747; PMID:3774547  
A:Accession: A25757  
A:Molecule type: DNA  
A:Residues: 1-1315 <FAI>  
A:Cross-references: GB:X06214; NID:g40773; PIDN:CAA29564.1; PID:g40774  
A:Experimental source: strain CN3911  
R:Fairweather, N.F.; Lyness, V.A.; Pickard, D.J.; Allen, G.; Thomson, R.O.  
J. Bacteriol. 165, 21-27, 1986  
A:Title: Cloning, nucleotide sequencing, and expression of tetanus toxin fragment C 1  
A:Reference number: A25194; MUID:86085672; PMID:3510187  
A:Accession: A25194  
A:Molecule type: DNA  
A:Residues: 743-1315 <FA2>  
A:Cross-references: GB:ML2739; NID:g144920; PIDN:AAA23282.1; PID:g144921  
A:Accession: B25194  
A:Molecule type: protein  
A:Residues: 865-894 <FA3>  
R:Matsuda, M.; Lei, D.L.; Sugimoto, N.; Ozutsumi, K.; Okabe, T.  
Infect. Immun. 57, 3588-3593, 1989  
A:Title: Isolation, purification, and characterization of fragment B, the NH-2-termin  
A:Reference number: A60759; MUID:90035436; PMID:2478476  
A:Accession: A60759  
A:Molecule type: protein  
A:Residues: 461-475 <MA>  
R:Demot, S.; Lanzavecchia, L.; Eisel, U.; Niemann, H.; Widmann, C.; Corradin, G.  
J. Immunol. 142, 394-402, 1989  
A:Title: Delineation of several DR-restricted tetanus toxin T cell epitopes.  
A:Reference number: JS0098; MUID:89093918; PMID:2463305  
A:Contents: annotation; epitope region  
R:Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta,  
Nature 359, 832-835, 1992  
A:Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteo  
A:Reference number: S27125; MUID:93063293; PMID:1331807  
A:Contents: annotation  
R:de Filippis, V.; Vangelista, L.; Schiavo, G.; Tonello, F.; Montecucco, C.  
Eur. J. Biochem. 229, 61-69, 1995  
A:Title: Structural studies on the zinc-endopeptidase light chain of tetanus neurotox  
A:Reference number: S69348; MUID:95262688; PMID:7744050  
A:Accession: S69348  
A:Molecule type: protein  
A:Residues: 2-31 <DEF>  
C:Comment: The source of this protein was an extrachromosomal plasmid.  
C:Comment: The precursor is cleaved by endogenous proteinase activity to form light ( dual chains are not toxic when separated). The amino end of the heavy chain (fragment C:Comment: Fragment B forms ion channels in a lipid bilayer. Fragment C binds to gang C:Comment: This potent neurotoxin binds to peripheral neuronal synapses, is internal presynaptic neurons. It inhibits neurotransmitter release by proteolytic cleavage of C:Function:  
A:Description: blocks neuroexocytosis via hydrolysis of a Gln-Phe peptide bond in syn A:Superfamily: tetanus toxin  
C:Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc F:2-457/Product: tentoxilysin light chain (fragment A) #status predicted <TTL> F:461-1315/Product: tentoxilysin heavy chain (fragment B.C) #status experimental <TTH F:461-864/Domain: channel forming (fragment B) #status predicted <XGB> F:865-1315/Domain: ganglioside binding (fragment C) #status predicted <TXC> F:233/237/Binding site: zinc (His) #status predicted F:234/Active site: Glu #status predicted

Query Match 17.89; Score 425; DB 1; Length 1315;  
Best Local Similarity 26.48; Pred. No. 8.5e-17;  
Matches 134; Conservative 90; Mismatches 169; Indels 114; Gaps 21;

QY 3 IPF----NIFSYNNLSKLDINIFYNNINDSKILSONKNTLVDTSGYNAEVEGVDV 58  
DB 859 IPFSYKSLDCVNDDEIDVI-----LKKSTILNDIINDIISGFSNVSIVTPDA 912  
QY 59 QLNPIFPDFKLGSGE-----DRGKVIYTONENIYVNSMFYSFISFWIRNK-WVS 110

Search completed: November 7, 2002, 14:46:49  
Job time : 53 secs





CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -!- MISCELLANEOUS: BOTULINUM TYPE C1 NEUROTOXIN IS SYNTHESIZED BY C  
 CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE  
 CC BACTERIOPHAGE.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X66433; CAA47060.1; -  
 CC EMBL; X72793; CAA51313.1; -  
 CC EMBL; X53751; CAA37780.1; -  
 CC EMBL; D90210; BAA14235.1; -  
 CC EMBL; X62389; CAA44263.1; -  
 CC PIR; S11291; S11291.  
 CC PIR; A35396; A35396.  
 CC PIR; A43503; A43503.  
 CC HSSP; P10845; 3BTA.  
 CC MEROPS; M27.002; -  
 CC InterPro; IPR000395; Bontoxilysin.  
 CC InterPro; IPR000130; Zn\_MiPeptidse.  
 CC Pfam; PF01742; Peptidase\_M27; 1.  
 CC PRINTS; PR00760; BONTOXILYSIN.  
 CC PRODOM; PD001963; Bontoxilysin; 1.  
 CC PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 FT INIT\_MET 0 0  
 FT CHAIN 1 448 BOTULINUM NEUROTOXIN C1, LIGHT-CHAIN.  
 FT CHAIN 449 1290 BOTULINUM NEUROTOXIN C1, HEAVY-CHAIN.  
 FT METAL 228 228 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 229 229 BY SIMILARITY.  
 FT METAL 232 232 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT DISULFID 436 452 INTERCHAIN (PROBABLE).  
 FT CONFLICT 84 84 P -> T (IN REF. 2).  
 SQ SEQUENCE 1290 AA; 148734 MW; 71FBE379F97129E8 CRC64;  
 Query Match 99.8%; Score 2386; DB 1; Length 1290;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-135;  
 Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC 2 TIFPFIYSYTNLSKDIINEYFNINDSKILSLQNKNTLVDTSGYNAEVSSEGDVQLN 61  
 DB 842 TIFPFIYSYTNLSKDIINEYFNINDSKILSLQNKNTLVDTSGYNAEVSSEGDVQLN 901  
 CC 62 PIFPFDKLGSGEDRGKVIYQTNENIVNYSMEYSFISFIRINKWVSLPGYTIIDSV 121  
 DB 902 PIFPFDKLGSGEDRGKVIYQTNENIVNYSMEYSFISFIRINKWVSLPGYTIIDSV 361  
 CC 122 KNSGWSGIISNLFVFLTKONESEOSINFSYDISNAPGYKNKFFVTVNNMGNMKI 181  
 DB 962 KNSGWSGIISNLFVFLTKONESEOSINFSYDISNAPGYKNKFFVTVNNMGNMKI 1021  
 CC 182 YNGKLIDITIKVKELTGINFSKTITFEINKIPDGLTSDSDINNMWIRDFYFAKELDG 241  
 DB 1022 YNGKLIDITIKVKELTGINFSKTITFEINKIPDGLTSDSDINNMWIRDFYFAKELDG 1081  
 CC 242 KDINILFNSQYTNVVDYWGNDLRYNKYEYVWVNDYLNRYMTVANSRQIVFNRNNDF 301  
 DB 1082 KDINILFNSQYTNVVDYWGNDLRYNKYEYVWVNDYLNRYMTVANSRQIVFNRNNDF 1141  
 CC 302 NQGYKIIKIRGNNDPRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDYVIG 361  
 DB 1142 NQGYKIIKIRGNNDPRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDYVIG 1201

QY 362 LRQQTQKIDNDNIIFQIPMNNYYIASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHN 421  
 DB 1202 LRQQTQKIDNDNIIFQIPMNNYYIASQIFKSNFNGENISGICSTGYRFRLLGGDWYRHN 1261  
 QY 422 YLPTVKQGNYSALLESTSTHWGFEVPE 450  
 DB 1262 YLPTVKQGNYSALLESTSTHWGFEVPE 1290  
 RESULT 2  
 ID BMD\_CLOBO STANDARD; PRT; 1276 AA.  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type D precursor (EC 3.4.24.69) (BONT/D)  
 DE (Bontoxilysin D).  
 GN BOTD.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBL\_TaxID=1491;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=BVD/-3;  
 RX MEDLINE=91016853; PubMed=2216736;  
 RA Binz T., Kurazono H., Popoff M.R., Eklund M.W., Sakaguchi G.,  
 RA Kozaki S., Krieglstein K., Henschen A., Gill D.M., Niemann H.;  
 RT "Nucleotide sequence of the gene encoding Clostridium botulinum  
 RT neurotoxin type D."; J. Vet. Med. Sci. 54:905-913(1992).  
 RL Nucleic Acids Res. 18:5556-5556(1990).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=CB16;  
 RX MEDLINE=93042276; PubMed=1420572;  
 RA Sunagawa H., Ohyama T., Watanabe T., Inoue K.;  
 RT "The complete amino acid sequence of the Clostridium botulinum type D  
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding  
 RT phage d-16 phi genome."; J. Vet. Med. Sci. 54:905-913(1992).  
 RN [3]  
 RC PARTIAL SEQUENCE.  
 RP STRAIN=D-SA, and D-1873;  
 RX MEDLINE=89339741; PubMed=2668193;  
 RA Morishita K., Syuto B., Kubo S., Oguma K.;  
 RT "Molecular diversity of neurotoxins from Clostridium botulinum type D  
 RT strains."; Infect. Immun. 57:2885-2891(1989).  
 RN [4]  
 RC IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94230352; PubMed=8175689;  
 RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,  
 RA Roques B., Fyke E.M., Suedhof T.C., Jahn R., Niemann H.;  
 RT "Cleavage of members of the synaptobrevin/VAMP family by types D and  
 RT F botulinum neurotoxins and tetanus toxin."; J. Biol. Chem. 269:12764-12772(1994).  
 CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CLEAVES THE 60-LYS-|-LEU-61 BOND OF  
 CC SYNAPTOSOMALIN-1 AND -2.  
 CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 CC detected action on small molecule substrates.  
 CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -!- MISCELLANEOUS: BOTULINUM TYPE D NEUROTOXIN IS SYNTHESIZED BY D  
 CC STRAIN OF CLOSTRIDIUM BOTULINUM WHICH CARRY THE APPROPRIATE  
 CC BACTERIOPHAGE.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X54254; CAA38175.1; -;  
 CC EMBL; S49407; AAB2444.1; -;  
 CC PIR; S11455; S11455.  
 CC DR HSSP; P10845; 3BTA.  
 CC DR MEROPS; M27.002; -;  
 CC DR InterPro; IPR000395; Bontoxilysin.  
 CC DR InterPro; IPR000130; Zn\_Mtpeptidse.  
 CC DR Pfam; PF01742; Peptidase\_M27; 1.  
 CC DR PRINTS; PR00760; BONTOXILYSIN.  
 CC DR ProDom; PD001963; Bontoxilysin; 1.  
 CC DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC DR CHAIN 1 442  
 CC FT BOTULINUM NEUROTOXIN D, LIGHT-CHAIN.  
 CC FT BOTULINUM NEUROTOXIN D, HEAVY-CHAIN.  
 CC FT ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT ACT\_SITE 230  
 CC FT METAL 233  
 CC FT DISULFID 437  
 CC FT VARIANT 15 16  
 CC FT VARIANT 17 18  
 CC FT VARIANT 452 452  
 CC FT VARIANT 457 457  
 CC FT VARIANT 457 457  
 CC FT VARIANT 457 457  
 CC FT VARIANT 457 457  
 CC FT VARIANT 489 489  
 CC FT VARIANT 544 644  
 CC FT VARIANT 1122 1122  
 CC SQ SEQUENCE 1276 AA; 146871 MW; C1EC50F46C8233E2 CRC64;  
 CC  
 CC Query Match 36.3%; Score 867.5; DB 1; Length 1276;  
 CC Best Local Similarity 43.2%; Pred. No. 2e-44;  
 CC Matches 195; Conservative 83; Mismatches 146; Indels 27; Gaps 11;  
 CC  
 CC 2 TIPFNIFSYTNNSLLKDIINEYFNNDKNSKILSLQNRKNTLVDTSGYNAEYSEGDYQLN 61  
 CC DB 839 TMPFNIFSYTNNSLLKDIINEYFNNDKNSKILSLQNRKNTLVDTSGYNAEYSEGDYQLN 898  
 CC QY 62 PIPFPFKLGSSGDEGRKVIYVQNSIVNYSVSEFSISFWIRLNKWSNLPGLPG-YTIIDS 120  
 CC DB 899 TITNDFKLGSSGD---KIIVNLNNLISAIYENSSVFWIKISKDUTNSHNYTIINS 955  
 CC QY 121 VKNSGWSGIISNLFVTLKQNEDEKSEISNFIYSIDISNAPGY-NKPFVTVTNMAGNM 179  
 CC DB 956 IEQNSGKLCIRNGNIEWILQDVNRKYSKLIIFYDESLSHTGYTNKFEVITINMGYM 1015  
 CC QY 180 KIYNGKLDITKVKELTGINSKTIPEINKIPDTGLTSDSNINKNWIRDFVIFAKEL 239  
 CC DB 1016 KLYNGELQSKQIEDLDEVKLDKTIYFVIGDENID-----ENQMLWIRDFNIFSKEL 1067  
 CC QY 240 DGKDINTLFSNLTQYTNVWDYNGNDRLYNKEYMYMVIDLYNRYMANSRQIVFVTRNRNN 299  
 CC DB 1068 SNEDINIVYEGQLRNVIKDYNGNPLKFTDEYIINDYVINDYVAPESNVLVLYQYDPS 1127  
 CC QY 300 DFNEGKIIIRKTRGNWNTDVRVGGDILYFDFTINKKAYNLFMKMKNWYADNH---STED 356  
 CC DB 1128 KLYTGNPITIKSYSDRNKPSRLINGDNIILHMLYNSRKYMIIRDYTDYIYVQGGECSSQC 1187  
 CC QY 357 IYALGLEQKQKINDNIIFIQIQPMNNTYYVYASQIFKPSNFGNENISSGCSI-GTYVRFRLGG 415

DB 1198 VYALKQSNLNGYIG-IFSINKNIVSKNKCQIF-SSFR-ENTMILLADYKWRFS--- 1241  
 QY 416 DWYRNLYLVPTVTKGKNVYASLLESTSTWGFV 446  
 DB 1242 --FKNAYTPVAVT--NYETKLLSTSSFWKEI 1268  
 CC  
 CC RESULT 3  
 CC BXG\_CLOBO STANDARD; PRT; 1296 AA.  
 CC AC Q60393;  
 CC DT 01-NOV-1997 (Rel. 35, Created)  
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE Botulinum neurotoxin type G precursor (EC 3.4.24.69) (Bont/G)  
 CC DE (Bontoxilysin G).  
 CC GN BORG.  
 CC OS Clostridium botulinum.  
 CC OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 CC OC Clostridium.  
 CC OX NCBI\_TaxID=1491;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX STRAIN=113 / 30;  
 CC RX MEDLINE=94092745; PubMed=8268233;  
 CC RA Campbell K., Collins M.D., East A.K.;  
 CC RT "Nucleotide sequence of the gene coding for Clostridium botulinum  
 CC (Clostridium argentinense) type G neurotoxin: genealogical comparison  
 CC with other clostridial neurotoxins";  
 CC RL Biochim. Biophys. Acta 1216:487-491(1993).  
 CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE.  
 CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 CC detected action on small molecule substrates.  
 CC -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X74162; CAA52275.1; -;  
 CC DR HSSP; P10845; 3BTA.  
 CC DR MEROPS; M27.002; -;  
 CC DR InterPro; IPR000395; Bontoxilysin.  
 CC DR InterPro; IPR000130; Zn\_Mtpeptidse.  
 CC DR Pfam; PF01742; Peptidase\_M27; 1.  
 CC DR PRINTS; PR00760; BONTOXILYSIN.  
 CC DR ProDom; PD001963; Bontoxilysin; 1.  
 CC DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 CC DR Neurotoxin; Hydrolase; Metalloprotease; Zinc.  
 CC INIT\_MET 0 0  
 CC BY SIMILARITY  
 CC FT CHAIN 1 441  
 CC FT BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.  
 CC FT BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.  
 CC FT ZINC (CATALYTIC) (BY SIMILARITY).  
 CC FT METAL 229 229  
 CC FT ACT\_SITE 230 230  
 CC FT METAL 233 233

```

FT   DISULFID      435    449          INTERCHAIN (PROBABLE).
SQ   SEQUENCE      1256 AA;  149013 MW;  DC8E47E15F665C31 CRC64;

Query Match       25.3%; Score 604.5; DB 1; Length 1296;
Best Local Similarity 32.0%; Pred. No. 9,1e-29;
Matches 156; Conservative 87; Mismatches 172; Indels 73; Gaps 16;

QY   2 TPIENIFSYTNKSLAKDIINEVFENNINDSKILSQRNEKNLVTDSYGNAEVSEEGDVOLN 61
     ||||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB   839 SPPFDLSLYTKTILLIQVFNMYISNISNAILSLSYRGGRLDISSGYGATMVGVSDYFVN 898

QY   62 PIPFPDFKLGGSGEDRGKVIVTVQENIYVNSWBSFSISFWIRINKWVEN-----LPQY 115
     ||||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB   899 DIGNGOFKLLNS--ENSNTAHQSKEVVYDMSFMDFNESINFEWETPKYNNDIQTYLQNEY 956

QY   116 TIIDSVKNSGSIGIGNFLVFTLKONEDSEQSINFSDISNNAPCY-NKWFEVTVTN 174
     ||||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB   957 TIIISKDNDSKWKSVIKGNRIIWLDVNAKSISIFEYSIKONISDYINKWFSTIIND 1016

QY   175 MGNMKIYYINGKLIDIITKKVELTGINTSFKITTFEINKIPTGTGLITSDDSNINMWIRDYFI 234
     :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB   1017 RLGNAMNIYYINGSKKSEKILNDRINSNDIDFKLINCTDTTF-----VWKDFNI 1069

QY   235 FAKELDGKDINILNSLOYTNVVDYWGNDLRYNKVEYNYNIDYLNRKYVANSROIVENT 294
     :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB   1069 FGRELNADEVSSLWIQTOSTNLTKDFWGNPLRYDTQYLFENOGMQNIYIKYFSKASMET 1128

QY   295 RRNNDFNE-----GKLIIRKI---RGWNTRVRGGDILYFDMTNNKAYNLF 341
     ||||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB   1129 APRTN-FNNAAINONLYGLGRFIRIKKASRN-NNDNIYREGDY-YLNI----- 1177

QY   342 MKNETWYADNHSTEDIYAIGREOTKTDINDNIIEFQIOPMNN--TYFYASQI----FKSNF 395
     :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB   1178 -----DNISDES-YRVYVLVNSKEIQTLF--LAPINDDPTEYDLVLIKKYEKITY 1226

QY   396 NGENIS-----GCISGTGTYRFLGGDWYRH-NYLVPYTKQGNYSALLESIS-----TH 442
     :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
DB   1227 NCOILCEKPTKTFGLFGIGKGVKDYGYVMDTYDNF--CISQWYLRRISENINKURLGCN 1284

QY   443 WGEFVPSVE 450
     |||||
DB   1285 WQFIPVDE 1292

RESULT 4
BXE_CLOBU
ID   BXE_CLOBU STANDARD; PRT; 1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, and ATCC 43755;
RA   MEDLINE=92181428; PubMed=1543481;
PM   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinum neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beiga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   J Biol Chem. 1992;267(1):107-113.
RM   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=B16340;
RA   MEDLINE=91237316; PubMed=2033376;
PM   Fujii N., Kimura K., Murakami T., Indah T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
```

QY 176 MGNMIVYINGKLDTIKVRELGTGINSFKTITEINKIPDTGLTSDSNINNNWIRDFYIF 235  
 Db 999 LGSOKUYINGNLDDKSIINLGNHIVSDNLEFKIVNCSTYRI-----GIRYFNIF 1049  
 QY 236 AKELGDKDINLNSLOYTNNVVKDYNGNDRNKEYYMWNI-----DYNRYWYANSROIV 291  
 Db 1050 DKELDETEIQTLYNENPNANILKDFWGNLYLDKEYILLNPKPNFNR--RUDSTLSI 1107  
 QY 292 FNTFRN---NNDNEGYKIIKRI-RGNTNDRVRGDDILYFDMTINKKAYNLFPMKNETM 347  
 Db 1108 NNIRSTILLANRLYSIGIKVQIWNSSNDNLVRKNDVOYINF-VASKTHLL-----PL 1161  
 QY 348 YADNHSTEDIYAGLREQTRQINDNINIFQIPMNNTYYVASOIFKSNFNGENISICSIG 407  
 Db 1162 YADTATINKEKTIKISSGNGRNFQVVM-----NSVGNCTMNFKN--NGNN-----IG 1208  
 QY 408 TYFR-----LGGDWYRHNLYVPVKOGNYSALLESSTH---WGFW 446  
 Db 1209 LLGFKADTVVASTWY-----YTHMDNINSNGFFWNI 1241

RESULT 5  
 BXAL\_CLOBO STANDARD; PRT; 1295 AA.  
 AC P10845; P18639; P01561;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)  
 DE (Bontolysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-  
 chain; Botulinum neurotoxin A, heavy-chain].  
 GN BOTA OR BNA OR ATX.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 CC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NTC 2916;  
 RX MEDLINE=90235864; PubMed=2185020;  
 RA Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,  
 RA Shone C.C., Atkinson T., Melling J., Minton N.P.;  
 RT "The complete amino acid sequence of the Clostridium botulinum type A  
 RT neurotoxin, deduced by nucleotide sequence analysis of the encoding  
 RT gene";  
 RL Eur. J. Biochem. 189:73-81(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=62A;  
 RX MEDLINE=90264400; PubMed=2160960;  
 RA Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;  
 RT "The complete sequence of botulinum neurotoxin type A and comparison  
 RT with other clostridial neurotoxins";  
 RL J. Biol. Chem. 265:9153-9158(1990).  
 RN [3]  
 RP SEQUENCE OF 1-65 FROM N.A.  
 RC STRAIN=62A;  
 RX MEDLINE=97016817; PubMed=8863443;  
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 RT "Organization and phylogenetic interrelationships of genes encoding  
 RT components of the botulinum toxin complex in proteolytic Clostridium  
 RT botulinum types A, B, and F: evidence of chimeric sequences in the  
 RT gene encoding the nontoxic nonhemagglutinin component";  
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 RN [4]  
 RP SEQUENCE OF 1-34 FROM N.A.  
 RC STRAIN=Hall;  
 RX MEDLINE=89350959; PubMed=2669749;  
 RA Betley M.J., Somers E., Dasgupta B.R.;  
 RT "Characterization of botulinum type A neurotoxin gene: delineation of  
 RT the N-terminal encoding region";  
 RL Biochem. Biophys. Res. Commun. 162:1388-1395(1989).

RN [5]  
 RP SEQUENCE OF 1-18 FROM N.A.  
 RC STRAIN=Type A NIH;  
 RX MEDLINE=96096783; PubMed=8521962;  
 RA Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;  
 RT "Molecular characterization of two forms of nontoxic-nonhemagglutinin  
 RT components of Clostridium botulinum type A progenitor toxins";  
 RL FEBS Lett. 376:41-44(1995).  
 RN [6]  
 RP SEQUENCE OF 1-16.  
 RX MEDLINE=84178501; PubMed=6370252;  
 RA Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;  
 RT "Partial amino acid sequence of the heavy and light chains of  
 RT botulinum neurotoxin type A";  
 RL Biochem. Biophys. Res. Commun. 119:900-904(1984).  
 RN [7]  
 RP SEQUENCE OF 1-46.  
 RX Dasgupta B.R., Foley J., Niece R.;  
 RT "Partial sequence of the light chain of botulinum neurotoxin type A";  
 RL Biochemistry 26:4162-4162(1987).  
 RN [8]  
 RP SEQUENCE OF 1-5 AND 444-456.  
 RX MEDLINE=91120847; PubMed=2126206;  
 RA Dasgupta B.R., Dekleva M.L.;  
 RT "Botulinum neurotoxin type A: sequence of amino acids at the  
 RT N-terminus and around the nicking site";  
 RL Biochimie 72:661-664(1990).  
 RN [9]  
 RP SEQUENCE OF 448-454 AND 872-895.  
 RX MEDLINE=89024662; PubMed=3178218;  
 RA Sartymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;  
 RT "Botulinum neurotoxin type A: cleavage of the heavy chain into two  
 RT halves and their partial sequences";  
 RL Arch. Biochem. Biophys. 266:142-151(1988).  
 RN [10]  
 RP SEQUENCE OF 448-482.  
 RX MEDLINE=85285016; PubMed=3896784;  
 RA Shone C.C., Hambleton P., Melling J.;  
 RT "Inactivation of Clostridium botulinum type A neurotoxin by trypsin  
 RT and purification of two tryptic fragments. Proteolytic action near  
 RT the COOH-terminus of the heavy subunit destroys toxin-binding  
 RT activity";  
 RL Eur. J. Biochem. 151:75-82(1985).  
 RN [11]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94063091; PubMed=8243676;  
 RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
 RA Benfenati F., Wilson M.C., Montecucco C.;  
 RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
 RT COOH-terminal peptide bonds";  
 RL FEBS Lett. 335:99-103(1993).  
 RN [12]  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=94124495; PubMed=8294407;  
 RA Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
 RA Jahn R., Niemann H.;  
 RT "Proteolysis of SNAP-25 by types E and A botulin neurotoxins";  
 RL J. Biol. Chem. 269:1617-1620(1994).  
 RN [13]  
 RP MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.  
 RX MEDLINE=21556941; PubMed=11700044;  
 RA Rigoni M., Caccini P., Johnson E.A., Montecucco C., Rossetto O.;  
 RT "Site-directed mutagenesis identifies active-site residues of the  
 RT light chain of botulinum neurotoxin type a";  
 RL Biochem. Biophys. Res. Commun. 288:1231-1237(2001).  
 RN [14]  
 RP X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).  
 RX MEDLINE=98455071; PubMed=9783750;  
 RA Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;  
 RT "Crystal structure of botulinum neurotoxin type A and implications  
 RT for toxicity";  
 RL Nat. Struct. Biol. 5:898-902(1998).  
 CC !- FUNCTION: Inhibits acetylcholine release. The botulinum toxin

[illegible][illegible]

RESULT 6	
BXE_CLOBO	STANDARD; PRT; 1250 AA.
ID BXE_CLOBO	
AC Q00496;	
DT 01-JUL-1993 (Rel. 26, Created)	
DT 01-JUL-1993 (Rel. 26, Last sequence update)	
DT 15-JUN-2002 (Rel. 41, Last annotation update)	
DE Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)	
DE (Bontoxilysin E).	
OS Clostridium botulinum.	
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;	
OC Clostridium.	
OX NCBI_TaxID=1491;	
[1]	
RN SEQUENCE FROM N.A.	
RC STRAIN=Beluga;	
RX MEDLINE=92181428; PubMed=1543481;	
PA Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;	
RT "Sequences of the botulinum neurotoxin E derived from Clostridium	
RT botulinum type E (strain Beluga) and Clostridium butyricum (strains	
RT ATCC 43181 and ATCC 43755).";	
RL Biochem. Biophys. Res. Commun. 183:107-113(1992).	
[2]	
RN SEQUENCE FROM N.A.	
RX MEDLINE=92174922; PubMed=1541280;	
RA Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;	
RT "The complete amino acid sequence of the Clostridium botulinum type-E	
RT neurotoxin, derived by nucleotide-sequence analysis of the encoding	
RT gene.";	
RL Eur. J. Biochem. 204:557-667(1992).	
[3]	
RN SEQUENCE OF 1-251 FROM N.A.	

RX MEDLINE=90264400; PubMed=21609560;  
RA Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;  
RT "The complete sequence of botulinum neurotoxin type A and comparison  
with other clostridial neurotoxins.";  
RL J. Biol. Chem. 265:9153-9158(1990).  
RN [4]  
RP SEQUENCE OF 1-13.  
RX MEDLINE=85197963; PubMed=3888113;  
RA Schmidt J.J., Sathiamoorthy V., Dasgupta B.R.;  
RT "Partial amino acid sequences of botulinum neurotoxins types B and  
E.";  
RN [5]  
RP SEQUENCE OF 419-426.  
RX MEDLINE=90344918; PubMed=21116911;  
RA Gimenez J.A., Dasgupta B.R.;  
RT "Botulinum neurotoxin type E fragmented with endoprotease Lys-C  
reveals the site trypsin nicks and homology with tetanus  
neurotoxin.";  
RL Biochimie 72:213-217(1990).  
RN [6]  
RP IDENTIFICATION OF SUBSTRATE.  
RX MEDLINE=94063091; PubMed=8243676;  
RA Schiavo G., Santucci A., Dasgupta B.R., Mehta P.P., Jontes J.,  
Bentzen F., Wilson M.C., Montecucco C.;  
RT "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct  
COOH-terminal peptide bonds.";  
RL FEBS Lett. 335:99-103(1993).  
RN [7]  
RP IDENTIFICATION OF SUBSTRATE.  
RX MEDLINE=94124495; PubMed=8294407;  
RA Binz T., Biasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,  
Jahn R., Niemann H.;  
RT "Proteolysis of SNAP-25 by types E and A botulinum neurotoxins.";  
RL J. Biol. Chem. 269:1617-1620(1994).  
CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
ENDOPETIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-  
181 BOND IN SNAP-25.  
CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
detected action on small molecule substrates.  
CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIANE CHANNEL  
FORMATION AND TOXIN BINDING, RESPECTIVELY.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
BOTULINUM NEUROTOXIN: TYPES A, B, C, D, E, F, AND G.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X62089; CAA43999.1; -;  
DR EMBL: X62683; CAA44558.1; -;  
DR PIR: A60027; A60027.  
DR PIR: B35294; B35294.  
DR PIR: JH0257; JH0257.  
DR PIR: S08575; S08575.  
DR PIR: S18111; S18111.  
DR PIR: S21178; S21178.  
DR HSP: P10845; 3BWA.  
DR MEROPS: M27.002; -;  
InterPro: IPR000395; Bontoxilysin.

DR InterPro: IPR000130; Zn\_Mtpeptidse.  
DR Pfam: PF01742; Peptidase\_M27; 1.  
DR PRINTS: PR00760; BONTOXILYSIN.  
DR ProDom: PD001963; Bontoxilysin; 1.  
DR PROSITE: PS00142; ZINC\_PROTEASE; 1.  
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
FT INIT\_MST 0  
FT CHAIN 1 421 BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.  
FT CHAIN 422 1250 BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.  
FT METAL 211 212 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 212 212 BY SIMILARITY.  
FT METAL 215 215 ZINC (CATALYTIC) (BY SIMILARITY).  
FT DISULFID 411 425 INTERCHAIN (PROBABLE).  
FT CONFLICT 176 176 R -> G (IN REF. 2).  
FT CONFLICT 197 197 C -> S (IN REF. 2 AND 3).  
FT CONFLICT 339 339 R -> A (IN REF. 2).  
FT CONFLICT 772 772 I -> L (IN REF. 2).  
FT CONFLICT 962 963 FE -> LQ (IN REF. 2).  
FT CONFLICT 966 966 R -> A (IN REF. 2).  
FT CONFLICT 1194 1194 N -> NN (IN REF. 2).  
SQ SEQUENCE 1250 AA; 143712 MW; D9FCE26DDA041EB4 CRC64;  
Query Match 24.4%; Score 583; DB 1; Length 1250;  
Best Local Similarity 31.9%; Pred. No. 1.7e-27;  
Matches 148; Conservative 88; Mismatches 166; Indels 62; Gaps 16;  
QY 2 TTPFNIFSYNNLSLLKDIINEYFNNDKSKILSNKRNKTLVDTSGYNAEVSSEGDYQLN 61  
DB 821 SIPEKLSVYDDKILISYENFFKRIKSSVNLNMYKNDKYVDTSGYDSNININGDVKY 880  
QY 62 PTFPDKLGSGEDRGKVIYVTONENIVNWSYFSFISFWIRI----NKWVSLPGYTI 117  
DB 881 PTNKNQF--GIYNDKLSSEVNSQNDYIIYDNKYNFSEFWIRIPNDYDNKIVNNNEYTI 938  
QY 118 IDSVK--NSNGSIGIISNLFVTLTKONEDSEQISNFSYDISNAPGY--YANSRQIVFN 175  
DB 939 INCMRDNSGKWSLNHNEIITWTFEDNGIKLAFNGANGISDYINKWFIWITNDR 998  
QY 176 MGNKIYINGKLIIDTKVKELTGINFSTIIFEINKIPDTGLTSDSDINNMWRDIYIF 235  
DB 999 LGDSKLYINGNIDOKSILNGLNIHVSNDILFKIVNCSYTRYI-----GIRYFNIF 1049  
QY 236 AKELGDKDNLNLFSLQTVNVYKVDGNDLRYNKEYVMNIDYLNRYM--YANSRQIVFN 293  
DB 1050 DKELDETEQTLISNEPWNILKDFGWYLLYDEYLLNVLKPNFIDRKDKSTLSINN 1109  
QY 294 TRRN---NNDFNEGKILIKRI--RGNTNDRVYRGDILYFDMT--NNKAYNLFMKNETMYA 349  
DB 1110 IRSTILLANRLYSGIKVKIQRVNNSSTNDNLVRKNDQYVINFVASKT--HLF----PLYA 1163  
QY 350 DNHSTEDIYALGRQTKDINDNIIFOQPMNNYVYASQIFKSNFNGENISGICISGTY 409  
DB 1164 DTATTNKTKIKISSGNRFNOVVYV-----NSVGNCTMFMKNN--NGNN-----IGLL 1210  
QY 410 RFR---LGGDNYRNYLPTVYKQGNYSALLSTSTH---WGTV 446  
DB 1211 GFKADIVVASTWY-----YTHMRDHTNSGCFWNFI 1241  
RESULT 7  
BXB\_CLOBO STANDARD; PRT; 1290 AA.  
ID BXB\_CLOBO  
AC P10844; P10843;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BONT/B)  
DE (Bontoxilysin B).  
GN BOTB.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1491;

RA RP SEQUENCE FROM N.A.  
 RX MEDLINE=92384550; PubMed=1514783;  
 RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,  
 RA Minton N.P.;  
 RT "Molecular cloning of the Clostridium botulinum structural gene  
 RT encoding the type B neurotoxin and determination of its entire  
 RT nucleotide sequence.";  
 RL Appl. Environ. Microbiol. 58:2345-2354(1992).  
 RN [2].  
 RN SEQUENCE OF 35-245 FROM N.A.  
 RC STRAIN=NCTC 7273;  
 RA Szabo B.A., Pemberton J.M., Desmarchellier P.M.;  
 RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.  
 RN [3].  
 RN SEQUENCE OF 633-993 FROM N.A.  
 RC STRAIN=NCTC 7273;  
 RX MEDLINE=94013372; PubMed=8408542;  
 RA Campbell K., East A.K., Collins M.D.;  
 RT "Gene probes for identification of the botulinum neurotoxin gene and  
 RT specific identification of neurotoxin types B, E, and F.";  
 RL J. Clin. Microbiol. 31:2255-2262(1993).  
 RN [4].  
 RN SEQUENCE OF 1-44 AND 441-466.  
 RC STRAIN=657;  
 RX MEDLINE=89000987; PubMed=3139097;  
 RA Dasgupta B.R., Datta A.;  
 RT "Botulinum neurotoxin type B (strain 657): partial sequence and  
 RT similarity with tetanus toxin.";  
 RL Biochimie 70:811-817(1988).  
 RN [5].  
 RN SEQUENCE OF 1-16 AND 441-458.  
 RC STRAIN=OKRA;  
 RX MEDLINE=85197963; PubMed=3886113;  
 RA Schmidt J.J., Sathyanarayanan V., Dasgupta B.R.;  
 RT "Partial amino acid sequences of botulinum neurotoxins types B and  
 RT E.";  
 RL Arch. Biochem. Biophys. 238:544-548(1985).  
 RN [6].  
 RN IDENTIFICATION AS ZINC-PROTEASE.  
 RX MEDLINE=93054694; PubMed=1429690;  
 RA Schlavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;  
 RT "Botulinum neurotoxins are zinc proteins.";  
 RL J. Biol. Chem. 267:23479-23483(1992).  
 RN [7].  
 RN IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE=93063293; PubMed=1331807;  
 RA Dasgupta B., Benfenati F., Poullain B., Rossetto O., de Laureto P.P.,  
 RA Schlavo G., Montecucco C.;  
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 RT by proteolytic cleavage of synaptobrevin.";  
 RL Nature 359:832-835(1992).  
 CC -1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-1-PHE-77 BOND OF  
 CC SYNAPTOSOMALIN-2.  
 CC -1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the  
 CC neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No  
 CC detected action on small molecule substrates.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A  
 CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,  
 CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIANE CHANNEL  
 CC FORMATION AND TOXIN BINDING, RESPECTIVELY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF  
 CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M81186; AAA23211.1; -;  
 DR EMBL; Z11934; CAA77991.1; -;  
 DR EMBL; X70817; CAA50148.1; -;  
 DR PIR; S07128; S07128;  
 DR PIR; S07155; S07155;  
 DR PIR; S08562; S08562;  
 DR PIR; S08573; S08573;  
 DR PIR; S08574; S08574;  
 DR PIR; A48940; A48940;  
 DR HSSP; P10845; 3BTA;  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.  
 DR InterPro; IPR000130; Zn\_MTPeptidase.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PR00760; BONTOXILYSIN.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 FT INIT\_MET 0  
 FT CHAIN 1 440 BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.  
 FT CHAIN 441 1290 BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.  
 FT METAL 229 229 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT\_SITE 230 230 BY SIMILARITY.  
 FT METAL 233 233 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 436 445 INTERCHAIN (PROBABLE).  
 FT DISULFID 29 29 T -> M (IN REF. 4).  
 FT CONFLICT 217 217 R -> G (IN REF. 2).  
 FT CONFLICT 224 224 A -> S (IN REF. 2).  
 FT CONFLICT 463 463 S -> R (IN REF. 4).  
 FT CONFLICT 463 463 S -> R (IN REF. 4).  
 SQ SEQUENCE 1290 AA; 150670 MW; D21746E2C024DF43 CRC64;  
 Query Match 24.3%; Score 581; DB 1; Length 1290;  
 Best Local Similarity 32.9%; Pred. No. 2.3e-27;  
 Matches 161; Conservative 82; Mismatches 168; Indels 78; Gaps 21;  
 QY 3 IPFNFSYNNSLKLDINEYNNINDSKILSONRKNLTVDTSGYNAEYSEGDVQLNP 62  
 DB 835 MPFOLSIYNTDILIEFMFNKYNSEILNLIILNRYKONLIDLSGYCAKVEYDGVDELN 894  
 QY 63 IFFPDFKLGSSEGRGVIVYTONENIVYNSVFSFISFWIRINKW---VSNL--PGYT 116  
 DB 895 --KNQFKLTSSA--NSKIRVTQNNLIIFNSVFLDFSVFWIRIPKYNKDCIQNYIENET 950  
 QY 117 IDSYYKNSGWSIGIISNLFVTLKQNEDESEQSINFSYDISNNAPGY-NKWFVFTVNNM 175  
 DB 951 IINCNNNSGWNKISIRGNRIITWTLIDINGKTKSVFFEYNIREDISEYINRWFVTTINN- 1009  
 QY 176 MGNKIYINGKLIDITIKVELTGINFSGKTIFFELINKPTDGLITSDSDNIN-NWIRDFYI 234  
 DB 1010 LNNAKIYINGKLENDWDIKDIREVANGELIFKLD-----GIDRTQFIWMYFSSI 1060  
 QY 235 FAKELDGKDINILFNSLQYITNVVKDYMGNDLRYNKEYMYNIDILNRYMYANS-----RQI 290  
 DB 1061 FNTLESQSNIEERYKIQSYSEYKDFWGNPLMYNKEYMFNAGNKNSYIKLKDKSPVGEI 1120  
 QY 291 VFNTERNN-----DFNEGYKIIIRKRGNT---NDTRVGGDILYFD-WTINKK--- 336  
 DB 1121 LTRSKYNSKYNINRYDLYIGKEFIIRK-XSNSQSINDIVRKEDYIYDFFLNENWVR 1179  
 QY 337 -AYNLFMKNETMYADNHSTEDIYAIGLRETKDINDNIIIFIQPMNNTYYASQ-IFKSN 394  
 DB 1180 YTYKYFKEE-----EKLFLAPISDSDEFYN---TIIKEYDEQPYSCQLLEK- 1226  
 QY 395 FNGENISICISGYRFLGG-----DWTNRHNYLVPTKQGNYSALLSESTST 441  
 DB 1227 -DEESTDEIGLIGHRFVESGIVPEYKDYFCISKWY-----LKEVKKRPYNLKL---GC 1277



QY 442 HWGFVPVSE 450  
 Db 1278 NWQFIPKDE 1286

## RESULT 8

EXA2\_CLOBO STANDARD; PRT; 1295 AA.  
 AC Q45894; P77780;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BONT/A)  
 DE (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-chain; Botulinum neurotoxin A, heavy-chain].  
 GN BOTA OR BNA OR ATX  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Kyoto-F;  
 RA MEDLINE=9413603; PubMed=8310180;  
 RA Willens A., East A.K., Lawson P.A., Collins M.D.;  
 RT "Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other clostridial neurotoxins.";  
 RL Res. Microbiol. 144:547-556(1993).  
 RN [2]  
 RP SEQUENCE OF 1-65 FROM N.A.  
 RC STRAIN=Kyoto-F;  
 RA MEDLINE=97016817; PubMed=8863443;  
 RA East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;  
 RT "Organization and phylogenetic interrelationships of genes encoding components of the botulinum toxin complex in proteolytic Clostridium botulinum types A, B, and F: evidence of chimeric sequences in the gene encoding the nontoxic nonhemagglutinin component.";  
 RL Int. J. Syst. Bacteriol. 46:1105-1112(1996).  
 CC -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin binds with high affinity to peripheral neuronal presynaptic membrane, is then internalized by receptor-mediated endocytosis. The C-terminus of the heavy chain (H) is responsible for the adherence of the toxin to the cell surface while the N-terminus mediates transport of the light chain from the endocytic vesicle to the cytosol. After translocation, the light chain (L) hydrolyzes the 197-Gln-1-Arg-198 bond in SNAP-25, thereby blocking neurotransmitter release. Inhibition of acetylcholine release results in flaccid paralysis, with frequent heart or respiratory failure (By similarity).  
 CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.  
 CC -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a heavy chain (H) (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MISCELLANEOUS: There are seven antigenically distinct forms of botulinum neurotoxin: Types A, B, C1, D, E, F, and G.  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 CC -----  
 DR EMBL; X73423; CAA51824.1; -;  
 DR EMBL; X87974; CAA61234.1; -;  
 DR HSP; P10845; 3BTA.  
 DR MEROPS; M27.002; -;  
 DR InterPro; IPR000395; Bontoxilysin.

DR InterPro; IPR000130; zn\_MTpeptdse.  
 DR Pfam; PF01742; Peptidase\_M27; 1.  
 DR ProDom; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; Hydrolase; Metalloprotease; Zinc.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
 FT INIT\_MET 0 0  
 FT CHAIN 1 447  
 FT CHAIN 448 1295  
 FT METAL 222 222  
 FT ACT\_SITE 223 223  
 FT METAL 226 226  
 FT DISULFID 429 453  
 FT DISULFID 1234 1279  
 FT TRANSMEM 626 646  
 FT TRANSMEM 655 675  
 FT SEQUENCE 1295 AA; 149279 MW; 5DA04A13D98D6372 CRC64;

Query Match 23.2%; Score 554.5; DB 1; Length 1295;  
 Best Local Similarity 30.8%; Pred. No. 8.6e-26;  
 Matches 147; Conservative 81; Mismatches 183; Indels 67; Gaps 12;

QY 3 IPFNIPSYTNNLLKDIINEYFNNINDSKILSLQNKNTLVDTSGYNAEVSEEGDVLNP 62  
 Db 848 IPQLSKYVDNKKLLSTFTTEYIKNIWTSILSVYKDDLDLSRYGAKINIGRVYDS 907  
 QY 63 IFPDFKLSSGDEGRKVIQVTONENIVNYSMEFSISFWIRNKVSNL---PGYTIID 119  
 Db 908 IDKNOIKLINESSTIEVL--KNAIVNYSMEVFSTFWIKPYFSKINLNNEYTIIN 965  
 QY 120 SVKNNSGWSGIISNPLVFTLKONEDSEOSINFSYDYSNNAPGY-NKAPFFVTVTNNMGN 178  
 Db 966 CIENNSGKVSINLYGELIITWLDQNKONIQRVVFKISQWYNSIDYINRWTFVITNNRLTK 1025  
 QY 179 MKIYINGKLDTIKVKELTGINFSTKITTEINKIPDTGLTSDSDNNINWIDFYFAKE 238  
 Db 1026 SKIYINGRLIDQKPSINLGNHASNKIMFKLDCRD-----PRRYIMIKYFNLFOKE 1077  
 QY 239 LDGKDINILNSLOQYINVKDYWGNDLYNKKEYMYNIDYLNRYVANS----- 287  
 Db 1078 LNEKEIKDLYDSOSNGILKDFWGNLYQDKPYIMJLNFDPKRYVDVNNIGRGWYILKG 1137  
 QY 288 --RQIVENTRRNNDFNEGKIIKIRGINTNDRVRGGDILYFDMTINNKAYNLFMKNE 345  
 Db 1138 PRGSVVTNITLNSLTVEGTFKIKKYGSGNEDNIVRNRDRIYINVVVKNKEYRL----- 1192  
 QY 346 TMDADNHESTEDIYGLRGQTKDINDINIFQI-----QPMNNTYYVASQIFKSNFNG 397  
 Db 1193 ---ATNASQAGVEKILSALEIPDVG--LSQVVVVKSKDDQGIKRN-----CKMNLQD 1240  
 QY 398 ENISGICISGTYRF-----RLGGDWYRHNLYLPTVKQNYASLLESTSTHWGFVPVSE 450  
 Db 1241 NNGNDIGFIFHLYDNIATKLIVASNWYN-----RQVGKAS--RTFGCSWEFIPVDD 1288

## RESULT 9

EXF\_CLOBO STANDARD; PRT; 1274 AA.  
 AC P30936;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BONT/F)  
 DE (Bontoxilysin F).  
 GN BOTF.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 23387;  
 RX MEDLINE=93012902; PubMed=1398040;  
 RA East A.K., Richardson P.T., Allaway D., Collins M.D.,

Roberts T.A., Thompson D.E.;  
"Sequence of the gene encoding type F neurotoxin of Clostridium botulinum";  
FEMS Microbiol. Lett. 75:225-230(1992).  
[2]  
SEQUENCE OF 1-64 FROM N.A.  
STRAIN-Hobbs FT10;  
MEDLINE=94297488; PubMed=7764998;  
RA East A.K., Collins M.D.;  
"Conserved structure of genes encoding components of botulinum neurotoxin complex M and the sequence of the gene coding for the nontoxic component in nonproteolytic Clostridium botulinum type F.";  
Curr. Microbiol. 29:69-77(1994).  
[3]  
SEQUENCE OF 634-1002 FROM N.A.  
MEDLINE=94013372; PubMed=8408542;  
RA Campbell K., East A.K., Collins M.D.;  
"Gene probes for identification of the botulin neurotoxin gene and specific identification of neurotoxin types B, E, and F.";  
J. Clin. Microbiol. 31:2255-2262(1993).  
[4]  
IDENTIFICATION OF SUBSTRATES.  
MEDLINE=94230352; PubMed=8175689;  
RA Yamasaki S., Baumeister A., Binz T., Blas J., Link E., Cornille F.,  
Romesaki B., Fyffe E.M., Suedhof T.C., Jahn R., Niemann H.;  
"Cleavage of members of the synaptobrevin/VAMP family by types D and F botulin neurotoxins and tetanus toxin.";  
J. Biol. Chem. 269:12784-12772(1994).  
-1- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-LYS-59 BOND OF SYNAPTOSREVINS-1 AND -2.  
-1- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No substrate action on small molecule substrates.  
-1- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.  
-1- CELLULAR LOCATION: Secreted.  
-1- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.  
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M92906; AAA23263.1; -  
DR EMBL; S73676; AAC60475.1; -  
DR EMBL; X70820; CAA50151.1; -  
DR EMBL; X70816; CAA50147.1; -  
DR HSSP; P10845; 3BFA.  
DR MEROPS; M27.002; -  
DR InterPro; IPR000395; Bontoxilysin.  
DR InterPro; IPR000130; Zn\_MTPeptidse.  
DR Pfam; PF01742; Peptidase\_M27; 1.  
DR PRINTS; PR00760; BONTOXILYSIN.  
DR ProDom; PD001963; Bontoxilysin; 1.  
DR PROSITE; PS001142; ZINC\_PROTEASE; 1.  
KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.  
FT CHAIN 1 436 BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.  
FT CHAIN 437 1274 BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.  
FT METAL 227 227 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT\_SITE 228 228 FT  
FT METAL 231 231 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 231 231

RT fragment C in *Escherichia coli*.  
 RL J. Bacteriol. 165:21-27(1988).  
 RN [4]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE-90201034; PubMed-2108021;  
 RA Krieglstein K., Henschel A., Weller U., Habermann E.;  
 RR "Arrangement of disulfide bridges and positions of sulphydryl groups  
 RT in tetanus toxin".  
 RL Eur. J. Biochem. 188:39-45(1990).  
 RN [5]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE-92037649; PubMed-1935979;  
 RA Krieglstein K.G., Henschel A.E., Weller U., Habermann E.;  
 RR "Limited proteolysis of tetanus toxin. Relation to activity and  
 RT identification of cleavage sites".  
 RL Eur. J. Biochem. 202:41-51(1991).  
 RN [6]  
 RP IDENTIFICATION AS ZINC-PROTEASE.  
 RX MEDLINE-93010948; PubMed-1396558;  
 RA Schiavo G., Poulain B., Rossetto O., Benfenati F., Tauc L.,  
 RR Montecucco C.;  
 RT "Tetanus toxin is a zinc protein and its inhibition of  
 RL neurotransmitter release and protease activity depend on zinc".  
 RN EMBO J. 11:3577-3583(1992).  
 RP IDENTIFICATION OF SUBSTRATE.  
 RX MEDLINE-93063293; PubMed-1331807;  
 RA Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,  
 RR Dasgupta B.R., Montecucco C.;  
 RT "Tetanus and botulinum-B neurotoxins block neurotransmitter release  
 RL by proteolytic cleavage of synaptobrevin".  
 RN Nature 359:832-835(1992).  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 874-1314.  
 RX MEDLINE-97475217; PubMed-9334741;  
 RA Umland T.C., Winger L.M., Swaminathan S., Furey W.F., Schmidt J.J.,  
 RR Sax M.;  
 RT "Structure of the receptor binding fragment HC of tetanus  
 RL neurotoxin".  
 CC Nat. Struct. Biol. 4:788-792(1997).  
 CC -1- FUNCTION: TETANUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER  
 CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED  
 CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD  
 CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT  
 CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC  
 CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 76-GLN-|-PHE-77  
 CC BOND OF SYNAPTOSOMAL-2.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 76-GLN-|-PHE-77 BOND IN  
 CC SYNAPTOSOMAL-2.  
 CC -1- SUBUNIT: THE PRECURSOR POLYPEPTIDE IS SUBSEQUENTLY CLEAVED TO  
 CC YIELD SUBCHAINS L AND H. THESE REMAIN LINKED BY A DISULFIDE BRIDGE  
 CC AND ARE NON-TOXIC AFTER SEPARATION.  
 CC -1- MISCELLANEOUS: THE C-TERMINAL OF THE HEAVY CHAIN BINDS TO  
 CC GANGLIOSIDE RECEPTORS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X04436; CAA28033.1; -  
 CC EMBL; M12739; AAA3282.1; -  
 CC EMBL; X06214; CAA29564.1; -  
 CC PIR; A25689; BTCLTN  
 CC PDB; 1AP9; 29-APR-98.  
 CC PDB; 1AGD; 14-OCT-98.  
 CC MEROPS; M27.001; -  
 CC InterPro; IPR000395; Bontoxilysin.  
 CC InterPro; IPR000130; Zn\_Mtpeptidse.

DR Pfam: PF01742; Peptidase\_M27; 1.  
 DR PRINTS; PRO0760; BONTOXILYSIN.  
 DR PRODOM; PD001963; Bontoxilysin; 1.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 KW Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; Plasmid;  
 KW 3D-structure.  
 FT INIT\_MET 0  
 FT CHAIN 1 456  
 FT CHAIN 457 1314  
 FT METAL 232 232  
 FT ACT\_SITE 233 233  
 FT METAL 236 236  
 FT TRANSMEM 226 246  
 FT TRANSMEM 669 689  
 FT DISULFID 438 466  
 FT DISULFID 1076 1092  
 SQ SEQUENCE 1314 AA; 150550 MW; 134C3657133EF81D CRC64;  
 Query Match 17.8%; Score 425; DB 1; Length 1314;  
 Best Local Similarity 26.4%; Pred. No. 4.5e-18;  
 Matches 134; Conservative 90; Mismatches 169; Indels 114; Gaps 21;  
 QY 3 IPP----NIFSTNNSLKDIINEYFNNINDSKILQNRKNTLVDTSGYNAEYSEGDV 58  
 DB 858 IPFSYKNDLCWVDNEEDIVI-----LKRSTILNLDINDINDISDGFNSSVITYPDA 911  
 QY 59 QLNPIPFDFKLGSGE-----DRGKIVIVTQENIVYNSMYESFSISFWIRNK-WVS 110  
 DB 912 QLVP-----GINKATHLVNSESSEVIVHKAMDIEYNDMNFNTVFWLVPKVSAS 963  
 QY 111 NLP-----GYTIDSVKNN-----SGWSGIISNLFVTLKQNEDESGSINFSYDINNA 160  
 DB 964 HLEQYGTNEYSISSMKHSLSGSGWSVSLKGNLTLTKDSAGEVQITFR-DLPDKF 1022  
 QY 161 PGV--NKWFFVTVNNMGMKIVYNGKLIDIKVKELTGINFSTIIFTEINKIPDTGLI 218  
 DB 1023 NAYLANKWFIITNDRLSSANLYNGVLMSAEITGLGAREDDNNITLKDR----- 1075  
 QY 219 TSQSDNNKWIIRDFYFAKELGKQINILFNSLQTVNVVKYWGNDLYNKEYYMNVI-- 276  
 DB 1076 -CENNQVYSIDKFRIFCALPKETKLYTSITFLRDFWGNPLRYDTEYLLIPVAS 1134  
 QY 277 -----DYL--NRYMVANSRQIVFNRRNNDNFEGYKIIKIRGNTN-DTRV 321  
 DB 1135 SSKDVLKNTIDYMLTNAPSVYNGKLIYRRLYN-----GLKFIKRYTPNEISFV 1189  
 QY 322 RGGDILYFDMTINKAY-----NLFMKNMTYADNHSTEDIYAGLREOTKDINDNI 373  
 DB 1190 KSGDFIKLYSVYNNNEHIVGPKDGNFNNLDRILRVGYNAPGIPLYKKMEAVK----- 1243  
 QY 374 IFQIQPMNNTYYVASQIFKSNFNGENISGICSGIYGRRLGSD-----WYRHHYL 423  
 DB 1244 -----LRDLKYSVOL--KLYDDKNAS-LGLVGHNGIIGNDPNDRILIASNRY-FNHL 1293  
 QY 424 VPIVKGNYASLLESTSTHWGFVPVSE 450  
 DB 1294 KDXI-----LGCDWYFVPTDE 1309  
 RESULT 11  
 EXEN\_CLOBO STANDARD; PRT; 1162 AA.  
 AC P46082;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Botulinum neurotoxin type E, nontoxic component.  
 GN ENT-120.  
 OS Clostridium botulinum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1491;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=Washington;
RX MEDLINE=93195515; PubMed=8450310;
RA Fujii N., Kimura K., Yokosawa N., Yashiki T., Tsuzuki K., Oguma K.;
RT "The complete nucleotide sequence of the gene encoding the nontoxic
component of Clostridium botulinum type E progenitor toxin.";
RL J. Gen. Microbiol. 139:79-86(1993).
CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12697; BAA02194.1;
DR InterPro: IPR000395; Bontotoxylisin.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOTOXILYSIN.
DR PRODOM; PD001963; Bontotoxylisin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136856 MW; 96468EDDAAE0F39D CRC64;
Query Match 9.6%; Score 230.5; DB 1; Length 1162;
Best Local Similarity 22.8%; Pred. No. 1.5e-06;
Matches 96; Conservative 74; Mismatches 158; Indels 93; Gaps 17;
QY 6 NIFSNTNSLLKDIINEYFNINNDSKILSLQNRKNTLV-DTSGYNAEVSSEGVDQLNPIF 64
DQ 783 NLFNLYTELLIKEQTSFY-----ELSLYAFQEQDNNVIGDTSGKNTLVE-----Y 827
QY 65 PDFKLGSSGDRGKVIYQNEIVY-----NSMYSFSSISFWIRINKWVSNLPGYTI- 117
DQ 828 PKDIGLVYGINNNAHLTGANQKFTNDYFENGLTNFSLYFWLR-----NLKQNTIK 881
QY 118 ---IDSVKNNSGWSIGIISNPLVFTLKQNEDESEOSINFSYDISNAPGYNKFFVT-N 173
DQ 882 SKLIGSKEDNCGWEIFYFENGLVFNIIIDSGNEKNIVLS-NISNKS-----WHYIVISIN 935
QY 174 NMGMNKIYINGKLIDTKVKELTGINFSKITTEINKIPDTGLTSDSNINNWIRDFY 233
DQ 936 RLKDQLLIFIDNVLVANEDEKEILNIYSS-----DIISLSDNNNVYIEGLS 982
QY 234 IFAKELDGKIDNILENSLYQYTNVYKDWGNDLRYNKYVWYVNYD-----LNRYMYANSRQ 289
DQ 983 VLNKTINSNEILTDYFSLDNNSYIRNDFEILQYRNYELFNVYFPEIAINKIEQNNIY 1042
QY 290 IVFNTFRNNDFNEGYKIIIRKIRGNTNDR-----VRGGDILYFDMTINNK 337
DQ 1043 LSIINNENLNFKPLKFKLL-----NTNPNKQYQKWDDEVIFSLDTEKYLDISTIN-- 1094
QY 338 YNLFMKNMNETMYADNHSTEDIYAIGLRQETKDINDNIIFQIPM--NNTYYVASOIFKSNF 395
DQ 1095 -----NRIQLVDNKNNAQIFIIN-----NDIFISNCLTLTYNNVNVYLS-IKNQDY 1139
QY 396 N 396
DQ 1140 N 1140

```

## RESULT 12

```

BXCN_CLOBU
ID BXCN_CLOBU STANDARD; PRT; 1162 AA.
AC Q06366;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Botulinum neurotoxin type E, nontoxic component.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

```

```

CC Clostridium.
CC NCBI_TaxID=1492;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BL6340;
RX MEDLINE=93360835; PubMed=8355622;
RA Fujii N., Kimura K., Yokosawa N., Oguma K., Yashiki T.,
RA Takashi K., Ohya T., Isogai E., Isogai H.;
RT "Similarity in nucleotide sequence of the gene encoding nontoxic
component of botulinum toxin produced by toxigenic Clostridium
botulinum strain BL6340 and Clostridium botulinum type E strain
Mashike.";
RL Microbiol. Immunol. 37:395-398(1993).
CC -!- FUNCTION: THE NONTXIC COMPONENT IS NECESSARY TO MAINTAIN
CC TOXICITY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D12739; BAA02231.1;
DR InterPro: IPR000395; Bontotoxylisin.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOTOXILYSIN.
DR PRODOM; PD001963; Bontotoxylisin; 1.
KW Neurotoxin.
SQ SEQUENCE 1162 AA; 136829 MW; C86E9BE932DA78E4 CRC64;
Query Match 9.6%; Score 230.5; DB 1; Length 1162;
Best Local Similarity 22.9%; Pred. No. 1.5e-06;
Matches 97; Conservative 76; Mismatches 152; Indels 99; Gaps 19;
QY 6 NIFSNTNSLLKDIINEYFNINNDSKILSLQNRKNTLV-DTSGYNAEVSSEGVDQLNPIF 64
DQ 783 NLFNLYTELLIKEQTSFY-----ELSLYAFQEQDNNVIGDTSGKNTLVE-----Y 827
QY 65 PDFKLGSSGDRGKVIYQNEIVY-----NSMYSFSSISFWIRINKWVSNLPGYTI- 117
DQ 828 PKDIGLVYGINNNAHLTGANQKFTNDYFENGLTNFSLYFWLR-----NLNQTNIK 881
QY 118 ---IDSVKNNSGWSIGIISNPLVFTLKQNEDESEOSINFSYDISNAPGYNKFFVT-N 173
DQ 882 SKLIGSKEDNCGWEIFYFENGLVFNIIIDSGNEKNIVLS-NISNKS-----WHYIVISIN 935
QY 174 NMGMNKIYINGKLIDTKVKELTGINFSKITTEINKIPDTGLTSDSNINNWIRDFY 233
DQ 936 RLKDQLLIFIDNVLVANEDEKEILNIYSS-----DIISLSDNNNVYIEGLS 982
QY 234 IFAKELDGKIDNILENSLYQYTNVYKDWGNDLRYNKYVWYVNYD-----LNRYMYANSRQ 289
DQ 983 VLNKTINSNEILTDYFSLDNNSYIRNDFEILQYRNYELFNVYFPEIAINKIEQNNN-- 1040
QY 290 IVFNTFRNNDFNEGYKIIIRKIRGNTNDR-----VRGGDILYFDMTIN 334
DQ 1041 -IYLSNNENLNFKPLKFKLL-----NTNPNKQYQKWDDEVIFSLDTEKYLDISID 1093
QY 335 NKAYNLFMKNMNETMYADNHSTEDIYAIGLRQETKDINDNIIFQIPM--NNTYYVASOIFK 392
DQ 1094 N-----NRIQLVDNKNNAQIFIIN-----NDIFISNCLTLTYNNVNVYLS-IKN 1136
QY 393 SNFN 396
DQ 1137 QDYN 1140

```

## RESULT 13

```

BXCN_CLOBU
ID BXCN_CLOBU STANDARD; PRT; 1196 AA.
AC P46081;

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 7, 2002, 14:40:06 ; Search time 81 Seconds  
(without alignments)  
1144.706 Million cell updates/sec

Title: US-09-910-186a-10  
Perfect score: 2391  
Sequence: 1 MTIFNFSYNNSLKDI.....NYASLESTHNGFVPSV 450

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2386	99.8	1291	2 Q93HT3	Q93ht3 clostridium
2	1824	76.3	1285	2 Q9LBR1	Q9lbr1 clostridium
3	1823	76.2	1285	2 Q45967	Q45967 clostridium
4	882.5	36.9	1280	2 Q9LBS7	Q9lbs7 clostridium
5	882.5	36.9	1280	2 Q45849	Q45849 clostridium
6	867.5	36.3	1275	12 Q9QNG7	Q9qng7 clostridium
7	610	25.5	1291	2 Q933K0	Q933k0 clostridium
8	604	25.3	1291	2 Q9ZAU8	Q9zau8 clostridium
9	602.5	25.2	1255	2 Q9FAR6	Q9far6 clostridium
10	602	25.2	1291	2 Q93G71	Q93g71 clostridium
11	594	24.8	1251	2 Q9K395	Q9k395 clostridium
12	593	24.8	1291	2 Q9Q077	Q9q077 clostridium
13	575	24.0	441	2 Q9X708	Q9x708 clostridium
14	569	23.8	1268	2 Q45851	Q45851 clostridium
15	564	23.6	1278	2 Q97236	Q97236 clostridium
16	527.5	22.1	1280	2 Q9ZAU5	Q9zau5 clostridium

17	428	17.9	451	2	Q9LAI3	Q9lai3 clostridium
18	428	17.9	1310	2	Q93N27	Q93n27 clostridium
19	314	13.1	361	2	Q45848	Q45848 clostridium
20	312	13.0	361	2	Q45846	Q45846 clostridium
21	281.5	11.8	367	2	Q45861	Q45861 clostridium
22	281.5	11.8	367	2	Q45862	Q45862 clostridium
23	258	10.8	1197	2	Q45888	Q45888 clostridium
24	239.5	10.0	1198	2	Q06018	Q06018 clostridium
25	236.5	9.9	1197	2	Q33871	Q33871 clostridium
26	236.5	9.9	1197	2	Q9ZAJ9	Q9zaj9 clostridium
27	234	9.8	1197	2	P71117	P71117 clostridium
28	231.5	9.7	1198	2	Q69277	Q69277 clostridium
29	230.5	9.6	1198	2	Q45893	Q45893 clostridium
30	216.5	9.1	1193	2	Q45880	Q45880 clostridium
31	215.5	9.0	1161	2	Q45891	Q45891 clostridium
32	215.5	9.0	1161	2	Q69276	Q69276 clostridium
33	213	8.9	1193	2	Q45914	Q45914 clostridium
34	213	8.9	1193	2	P71107	P71107 clostridium
35	208	8.7	1162	2	Q9ZAJ6	Q9zaj6 clostridium
36	200	8.4	1196	2	Q9LBS8	Q9lbs8 clostridium
37	199	8.3	1165	2	Q45844	Q45844 clostridium
38	199	8.3	1165	2	Q45887	Q45887 clostridium
39	199	8.3	1196	2	Q45916	Q45916 clostridium
40	199	8.3	1196	2	Q93HT4	Q93ht4 clostridium
41	199	8.3	1196	9	Q38197	Q38197 clostridium
42	197	8.2	1163	2	Q45850	Q45850 clostridium
43	197	8.2	1196	2	Q9LBR2	Q9lbr2 clostridium
44	197	8.2	1196	2	Q53550	Q53550 clostridium
45	197	8.2	1196	9	Q9ZX77	Q9zx77 clostridium

#### ALIGNMENTS

RESULT 1  
Q93HT3 ID Q93HT3 PRELIMINARY; PRT; 1291 AA.  
AC Q93HT3;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE Neurotoxin.  
GN Nr.  
OS Clostridium botulinum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1491;  
[1]  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN-C-VOICHI;  
RX MEDLINE=21534265; PubMed=11676492;  
RA Sagane Y., Kouguchi H., Watanabe T., Sunagawa H., Inoue K.,  
RA Fujinaga Y., Oguma K., Ohyama T.;  
RT "Role of C-Terminal Region of HA-33 Component of Botulinum Toxin in  
RT Hemagglutination";  
RL Biochem. Biophys. Res. Commun. 288:650-657(2001).  
DR EMBL; AB061780; BABY1749.1; -;  
DR InterPro; IPR000395; Bontoxilysin.  
DR InterPro; IPR000130; Zn\_Mtpeptdse.  
DR Pfam; PF01742; Peptidase\_M27; 1.  
DR ProDom; PD001963; Bontoxilysin; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN1.  
SQ SEQUENCE 1291 AA; 148869 MW; 4A21DB35B8743CF8 CRC64;  
Query Match 99.8%; Score 2386; DB 2; Length 1291;  
Best Local Similarity 100.0%; Pred. No. 5.7e-117;  
Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 TIFNFSYNNSLKDIINEYFNNDISKILSQNRKNTLVDTSGYNAEVSSEGDVQLN 61  
Db 843 TIFNFSYNNSLKDIINEYFNNDISKILSQNRKNTLVDTSGYNAEVSSEGDVQLN 902  
QY 62 PIFPFDFKLGSSGDEGRGVIVTQNEINIVNSMYSFISFWIRNKWVSNLPGYTIIDS 121

```

Db 903 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRNKWSNLPGYTIIDSV 962
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYNKWFVTVTNMGMNMI 181
Db 963 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYNKWFVTVTNMGMNMI 1022
Qy 182 YINGKLIDTIVKELTGINFSGKITTEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 241
Db 1023 YINGKLIDTIVKELTGINFSGKITTEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 1082
Qy 242 KDINILNSLOYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYANRSQIVFTRNNNDP 301
Db 1083 KDINILNSLOYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYANRSQIVFTRNNNDP 1142
Qy 302 NEGKIIIRKIRGTNTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 361
Db 1143 NEGKIIIRKIRGTNTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1202
Qy 362 LREQTKDINDNIIFQIOPMNTYYASQIFKSNFNGENISGICSTGYRFRJGGDWYRHN 421
Db 1203 LREQTKDINDNIIFQIOPMNTYYASQIFKSNFNGENISGICSTGYRFRJGGDWYRHN 1262
Qy 422 YLVPTVKQGNVYASLESTSTHWGFVPVSE 450
Db 1263 YLVPTVKQGNVYASLESTSTHWGFVPVSE 1291

RESULT 2
Q9LBR1
ID Q9LBR1 PRELIMINARY; PRT: 1285 AA.
AC Q9LBR1
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D-4947;
RA Sagene Y., Watanabe T., Kouguchi H., Yamamoto T., Takizawa J.,
RA Kawabe T., Murakami F., Muroga A., Nakatsuka M., Ohyama T.;
RT "Characterization of the progenitor Toxin Components Produced by
RT Clostridium botulinum Type D Strain 4947".
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB037920; BAA90661.1; -
DR HSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1285 AA; 147352 MW; B63AFA487D570680 CRC64;

Query Match 76.3%; Score 1824; DB 2; Length 1285;
Best Local Similarity 77.9%; Pred. No. 1.3e-87;
Matches 352; Conservative 36; Mismatches 56; Indels 8; Gaps 4;

Qy 2 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61
Db 839 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 898
Qy 62 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRNKWSNLPGYTIIDSV 121
Db 899 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRNKWSNLPGYTIIDSV 958
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYNKWFVTVTNMGMNMI 181
Db 963 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYNKWFVTVTNMGMNMI 1022

```

```

Db 959 KNSGWSIGIISNLFVTLKQNESEQDINFSYDISKNAAGYNKWFVTVTNMGMNMI 1018
Qy 182 YINGKLIDTIVKELTGINFSGKITTEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDTIVKELTGINFSGKITTEINKIPDTGLTSDSDINNMWIRDFYIFAKELDG 1078
Qy 242 KDINILNSLOYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYANRSQIVFTRNNNDP 301
Db 1079 KDINILNSLOYTNVVDYWGNDLRYNKEYMYNIDYLNRYMYANRSQIVFTRNNNDP 1138
Qy 302 NEGKIIIRKIRGTNTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 360
Db 1139 NEGKIIIRKIRGTNTRVRGGDILYFDMTINNKAYNLFMKNETMYADNHSTEDIYAI 1193
Qy 361 GLRQTKD-INDNIIFQIOPMNTYYASQIFKSNFNGENISGICSTGYRFRJGGDWY 418
Db 1194 GALDQPMDETRKYGSTFIQPCNTFDYASQLFLSSNATNRLGILSISGYSFKLGDDYWF 1253
Qy 419 RHNLYPTVKQGNVYASLESTSTHWGFVPVSE 450
Db 1254 NHEYLPVIRKIEHYASLESTSTHWGFVPASE 1285

RESULT 3
Q45967
ID Q45967 PRELIMINARY; PRT: 1285 AA.
AC Q45967
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Neurotoxin CONSISTING of botulinum neurotoxin D and C1.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TYPE D;
RX MEDLINE=96283801; PubMed=8679691;
RA Moriishi K., Koura M., Abe N., Fujii N., Fujinaga Y., Inoue K.,
RA Ogumada K.;
RT "Mosaic structures of neurotoxins produced from Clostridium botulinum
RT types C and D organisms";
RL Biochim. Biophys. Acta 1307:123-126(1996).
DR EMBL; D38442; BAA07477.1; -
DR HSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Neurotoxin.
SQ SEQUENCE 1285 AA; 147365 MW; EDE98E4EAC6413 CRC64;

Query Match 76.2%; Score 1823; DB 2; Length 1285;
Best Local Similarity 77.7%; Pred. No. 1.5e-87;
Matches 351; Conservative 37; Mismatches 56; Indels 8; Gaps 4;

Qy 2 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 61
Db 839 TIPNIFSYTNNSLLKDIINEYFNNDKILSLQNRKNTLVDTSGYNAEVSSEGVDQLN 898
Qy 62 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRNKWSNLPGYTIIDSV 121
Db 899 PIFPFDFKLGSSGDRGKVIIVTQENIVYNSMYESFSISFWIRNKWSNLPGYTIIDSV 958
Qy 122 KNSGWSIGIISNLFVTLKQNEDESEOSINFSYDISNAPGYNKWFVTVTNMGMNMI 181
Db 959 KNSGWSIGIISNLFVTLKQNESEQDINFSYDISKNAAGYNKWFVTVTNMGMNMI 1018

```



```

QY 182 YINGKLIDIKVKELTGINFSTIIFKINPKIPDTGLTSDSDNINMWIRDFYIFAKELDG 241
Db 1019 YINGKLIDIKVKELTGINFSTIIFQNKIPNTGLTSDSDNINMWIRDFYIFAKELDD 1078
QY 242 KQINILFNSLQTNVVKDYWGNDLRVKNKEYYMWVNDYLNRYMYANSROIVFNTRENN 301
Db 1079 KQINILFNSLQTNVVKDYWGNDLRVKNKEYYMWVNDYLNRYMYANSROIVFNTRENN 1138
QY 302 NGEYKIIIRKIRGNTNDTRVRGDDILYFDMTINNKAYLNFMKNTMYADNHST-EDIYAI 360
Db 1139 NGEYKIIIRKIRGNTNDTRVRGDDILYFDMTINNKAYLNFMKNTMYADNHST-EDIYAI 1193
QY 361 GURETKD-INDNIIFOQPMNNTVYASQIFKSNFNGENISGICSTGYFRIGSD-WY 418
Db 1194 GALDQPMDEIRKYGSFIQPCNTFYASQIFKSNFNGENISGICSTGYFRIGSD-WY 1253
QY 419 RNYLVPTVKQNYASLLESTSTHGWFPVSE 450
Db 1254 NHEYLIPVIKIEHYASLLESTSTHGWFPVASE 1285

RESULT 4
Q9LBS7 PRELIMINARY; PRO; 1280 AA.
AC Q9LBS7:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Neurotoxin.
GN NT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-6814;
RA Sagane Y., Watanabe T., Kouguichi H., Yamamoto T., Kawabe T.,
RA Murakami F., Nakatsuka M., Ohyama T.;
RT "Organization of Gene Encoding Components of the Botulinum Progenitor
RT Toxin in Clostridium botulinum Type C Strain 6814: Evidence of
RT Chimeric Sequence in the Gene Encoding Each Component."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB037166; BAA089713.1;
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1280 AA; 147757 MW; DEL24FFB6F68450B CRC64;

Query Match
Best Local Similarity 36.9%; Score 882.5; DB 2; Length 1280;
Matches 198; Conservative 78; Mismatches 148; Indels 27; Gaps 10;

QY 2 TTPFNIFSYNNSLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEYSEGDVQLN 61
Db 843 TTPFNIFSYNNSLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEYSEGDVQLN 902
QY 62 PTFPFDKLGSGEDRGKVIYTONENIYVNSMYESFSISFWIRNKVSNLPG-YTIIDS 120
Db 903 TIYTNDFKLSSGD---KIIVNLNNILYSAIYENSSVFWIKSKDLTNSHNEYTIINS 959
QY 121 VKNNSGWSIGTISNFLTFTLTKONEDSEOSINFSYDISNAPGY-NKWFVYVTNNMGMN 179
Db 960 IKQNSGWLKICRANGIEWLQDINRKYSLFDYSESLSHGTGYNKFWFTITNNINGYM 1019
QY 180 KYINGKLIDITKVKELGINFSTIIFKINPKIPDTGLTSDSDNINMWIRDFYIFAKEL 239
Db 1020 KYINGELKQSERIEDLNEVKLDKTIIVFGIDENID-----ENQMLWIRDFNIFSKEL 1071

```

```

QY 240 DGKDINILFNSLQTNVVKDYWGNDLRVKNKEYYMWVNDYLNRYMYANSROIVFNTRENN 299
Db 1072 SNEDINIVFEGOILRNVDYWGNDLRVKNKEYYMWVNDYLNRYMYANSROIVFNTRENN 1131
QY 300 DFNEGYKIIIRKIRGNTNDTRVRGDDILYFDMTINNKAYLNFMKNTMYA---DNHSTED 356
Db 1132 KLYTGNPITIKSVSDKNPYSRILGNDINMFHMLYNSGKIYMIROTDIYATEGRECSKNC 1191
QY 357 IYAIGLRQETKDINDNIIFOQPMNNTVYASQIFKSNFNGENISGICST-GTYRFRLLGG 415
Db 1192 VYALQSNLNGYGG-IFSINIVSQNKYCSQIFSSFM---KNTMLLADYKPMRFSF-- 1246
QY 416 DWYRHNLYVPTVKQNYASLLESTSTHGWGFV 446
Db 1247 ---ENAYTP-VAVTNYETKLLSTSTSFWKFI 1272

RESULT 5
Q45849 PRELIMINARY; PRT; 1280 AA.
AC Q45849;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Neurotoxin.
OS Clostridium botulinum C.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=36828;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6813; PubMed=8593068;
RX MEDLINE=96156810; PubMed=8593068;
RA Moriishi K., Koura M., Fujii N., Fujinaga Y., Inoue K., Syuto B.,
RA Oguma K.;
RT "Molecular cloning of the gene encoding the mosaic neurotoxin,
RT composed of parts of botulinum neurotoxin types C1 and D, and PCR
RT detection of this gene from Clostridium botulinum type C organisms."
RL Appl. Environ. Microbiol. 62:662-667(1996).
DR EMBL; D49440; BAA08418.1;
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_Mtpeptidse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1280 AA; 147815 MW; 74F813B228B8C989 CRC64;

Query Match
Best Local Similarity 36.9%; Score 882.5; DB 2; Length 1280;
Matches 198; Conservative 78; Mismatches 148; Indels 27; Gaps 10;

QY 2 TTPFNIFSYNNSLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEYSEGDVQLN 61
Db 843 TTPFNIFSYNNSLKDIINEYFNINDSKILSQNRKNTLVDTSGYNAEYSEGDVQLN 902
QY 62 PTFPFDKLGSGEDRGKVIYTONENIYVNSMYESFSISFWIRNKVSNLPG-YTIIDS 120
Db 903 TIYTNDFKLSSGD---KIIVNLNNILYSAIYENSSVFWIKSKDLTNSHNEYTIINS 959
QY 121 VKNNSGWSIGTISNFLTFTLTKONEDSEOSINFSYDISNAPGY-NKWFVYVTNNMGMN 179
Db 960 IKQNSGWLKICRANGIEWLQDINRKYSLFDYSESLSHGTGYNKFWFTITNNINGYM 1019
QY 180 KYINGKLIDITKVKELGINFSTIIFKINPKIPDTGLTSDSDNINMWIRDFYIFAKEL 239
Db 1020 KYINGELKQSERIEDLNEVKLDKTIIVFGIDENID-----ENQMLWIRDFNIFSKEL 1071
QY 240 DGKDINILFNSLQTNVVKDYWGNDLRVKNKEYYMWVNDYLNRYMYANSROIVFNTRENN 299

```

```
Db 1072 SNEDINIVEGQILRNWIKDYWGPNLKFDEYXYIINDNYIDRYIAPKSNILVLVQVDRS 1131
Qy 300 DFNEGYKIIIKRIGRNDTRVRGGRDILYFDMTNNKAYNLFMKNETMYA---DNHSTED 356
Db 1132 KLYTGNPTITKSVDDKPNYSRLNGDNIMHMLYNSKYMIRDTDTIYVIGRECSKNC 1191
Qy 357 IYATGLRQTKDINDNIIFQIQPMNNTYYVQSIKFNNGENISGICSI-GTYRFLGG 415
Db 1192 VYALKLQNLGNYGIG-IFSINKIVSONKYSQIFSSFM--KNTMLLADIYKPRFSE-- 1246
Qy 416 DWYRHNYLPTVKQGNVYASLLESTSTHWGFV 446
Db 1247 ---ENAYTP-VATNVTETKLLSTSSFWKFI 1272

RESULT 6
Q9QTG7 PRELIMINARY; PRT; 1275 AA.
AC Q9QTG7;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE NTX (Fragment).
GN Clostridium botulinum D bacteriophage.
OC Viruses.
OX NCBI_TaxID=29342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1873;
RA MEDLINE=99017546; PubMed=9802560;
RA Nakajima H., Inoue K., Ikeda T., Fujinaga Y., Sunagawa H., Takeshi K.,
RA Ohyama T., Watanabe T., Inoue K., Oguma K.;
RT "Molecular composition of the 16S toxin produced by a Clostridium
RT botulinum type D strain, 1873."
RL Microbiol. Immunol. 42:599-605(1998).
DR EMBL; AB012112; BAA75084.1; -
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; Bontoxilysin.
DR PRODOM; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
FT NON_TER 1275 1275
SQ SEQUENCE 1275 AA; 146742 MW; 3C50F46C8233E2D6 CRC64;

Query Match 36.3%; Score 867.5; DB 12; Length 1275;
Best Local Similarity 43.2%; Pred. No. 1.3e-37;
Matches 195; Conservative 83; Mismatches 146; Indels 27; Gaps 11;

Qy 2 TIPNIFSYTNNLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLN 61
Db 839 TAMPNIFSYTNNLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLN 898
Qy 62 PIFPFDKLGSSGDRGVIVTQENIVYNSMYESPISFWIRKWNLP-YYTIIDS 120
Db 899 TYTNDKFLSSGD--KLIYVNNLLYSAIYENSVSFWIKSKDLTNSHNEYTIIS 955
Qy 121 VKNNSGWSIGIISNLFVTLKQNEDEQSINFSYDVSNNAPGY-NKWFVFTVNNMGNM 179
Db 956 LEQNSGKLCIRNGNIEWLODVRNKYSLLFDYSESLGTGYNKWFVFTVNNMGNM 1015
Qy 180 KIYNGKLIDIKVKEGTGNFSTIIFENKIPDTGLTSDSDNIMMTRPIFAKEL 239
Db 1016 KLYINGELQSKQIEDLEVKLDRTIVFGDENID-----ENQMLTRDFNIFSKEL 1067
Qy 240 DGKDNILFSLQVTNVYKDWGNDLYNKEYYMNIDYLNRYMANSRQIVFTRNNN 299
Db 1068 SNEDINIVEGQILRNWIKDYWGPNLKFDEYXYIINDNYIDRYIAPKSNILVLVQVDRS 1127
Qy 300 DFNEGYKIIIKRIGRNDTRVRGGRDILYFDMTNNKAYNLFMKNETMYA---DNHSTED 356
```

```
Db 1128 KLYTGNPTITKSVDDKPNYSRLNGDNIMHMLYNSKYMIRDTDTIYVIGRECSQNC 1187
Qy 357 IYATGLRQTKDINDNIIFQIQPMNNTYYVQSIKFNNGENISGICSI-GTYRFLGG 415
Db 1188 VYALKLQNLGNYGIG-IFSINKIVSONKYSQIF--SSFR-ENTMLLADIYKPRFSE--- 1241
Qy 416 DWYRHNYLPTVKQGNVYASLLESTSTHWGFV 446
Db 1242 --PKNAVTPAVT--NYETKLLSTSSFWKFI 1268

RESULT 7
Q933KO PRELIMINARY; PRT; 1291 AA.
AC Q933KO;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Type B cryptic neurotoxin.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=593, AND 588;
RA Kirma N., Ferreira J.L., Baumstark B.R.;
RT "Characterization of six type A strains of Clostridium botulinum that
RT contain type B toxin gene sequences."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF300466; AAL11499.1; -.
DR EMBL; AF300465; AAL11498.1; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTPeptidase.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRODOM; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Neurotoxin.
SQ SEQUENCE 1291 AA; 150843 MW; 7AC1737B0FA5A151 CRC64;

Query Match 25.5%; Score 610; DB 2; Length 1291;
Best Local Similarity 33.9%; Pred. No. 3.7e-24;
Matches 168; Conservative 84; Mismatches 154; Indels 90; Gaps 20;

Qy 2 TIPNIFSYTNNLLKDIINEYFNNDKSLQNRKNTLVDTSGYNAEVSSEGDVQLN 61
Db 835 SIPFDLSTYNTNLTILIEFPKNYSDILNLIILNRYDNKLDLSYGAKEVYDGVKLN 894
Qy 62 PIFPFDKLGSSGDRGVIVTQENIVYNSMYESPISFWIRKWNLP-YYTIIDS 115
Db 895 D--KNQPKLTSSA--NSKIRVTQNIIFNSMFLDPSVFWIRIPKRYKNDGIONTHNEY 950
Qy 116 TIIDSVKNNSGWSIGIISNLFVTLKQNEDEQSINFSYDVSNNAPGY-NKWFVFTVNN 174
Db 951 TIINCNNKSNKSIIRGNMIWTLDINGKIKSVFEYSIKEDISEYINRWFVFTVNN 1010
Qy 175 MGNKMYIYNGKLIDTKVKELGINFPSKTIIFENKIPDTGLTSDSDNIN---MWIR 230
Db 1011 -SDNAKYIYNGKLESHIDIRDIANDIETKLD-----GNIDRTOFIWK 1057
Qy 231 DFYIFAKELGKDNILFSLQVTNVYKDWGNDLYNKEYYMNIDYLNRYM---AN 286
Db 1058 YFSFNTLQSQNIIEYKIQSYSEVLKDFWGNPLMYNKEYYMFNAGNKSYSIKLKQSS 1117
Qy 287 SRQIVFTRNNN-----DFNEGKYLIIKIRGNT---NDTRVRGGDILYFDMTINN 336
Db 1118 VGEILLTRSKYQNSKYINRYDLVIGKFIIR-KNSQSQINDDIVRKEDYIYLDFFNLNQ 1176
Qy 337 AYNLFMKNETMYADNHDSTEDIYALGRECTK---DINDNIIF---QIQPMNNTYYVAS 388
Db 1177 EWRVIM-----YKFKKEEKLFLAPISDSDEFYNTIQIKEYDEQPTVSC 1221
```





RA	Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,	
RA	Karasawa T.;	
RA	"C. butyricum (KZ 1890) gene for type E botulinum toxin.";	
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.	
RL	[10]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=KZ 1891;	
RC	Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,	
RA	Karasawa T.;	
RA	"C. butyricum (KZ 1891) gene for type E botulinum toxin.";	
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.	
RL	[11]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=LCL 063;	
RC	Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,	
RA	Karasawa T.;	
RA	"C. butyricum (LCL 063) gene for type E botulinum toxin.";	
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AB037714; BAB03522.1; -	
DR	EMBL; AB037704; BAB03512.1; -	
DR	EMBL; AB037705; BAB03513.1; -	
DR	EMBL; AB037706; BAB03514.1; -	
DR	EMBL; AB037707; BAB03515.1; -	
DR	EMBL; AB037708; BAB03516.1; -	
DR	EMBL; AB037709; BAB03517.1; -	
DR	EMBL; AB037710; BAB03518.1; -	
DR	EMBL; AB037711; BAB03519.1; -	
DR	EMBL; AB037712; BAB03520.1; -	
DR	EMBL; AB037713; BAB03521.1; -	
DR	HSSP; P10845; 3BTA. -	
DR	MEROPS; M27.002; -	
DR	InterPro; IPR000395; Bontoxilysin.	
DR	InterPro; IPR000130; Zn.Mtpeptidse.	
DR	Pfam; PF01742; Peptidase_M27; 1.	
DR	PRINTS; PR00760; BONTOXILYSIN.	
DR	ProDom; PD001963; Bontoxilysin; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.	
SQ	SEQUENCE 1251 AA; 143571 MW; 2021F4E427070296 CRC64;	
Query Match 24.8%; Score 594; DB 2; Length 1251;		
Best Local Similarity 32.2%; Pred. No. 2.5e-23;		
Matches 150; Conservative 94; Mismatches 156; Indels 66; Gaps		
QY	2 TIFNIFSYTNSLLKDIINEYFNINDSKILSQNRKNTLVDSGYNAEVSSEGQVQLN 61	
Db	822 SIPFKLSSYTDKKILISYENKPFKRIKSSVLLNMYKNDKYVDATSGYDSNINGEIFY 881	
QY	62 PIFPFDFKLGSGEDRGKVIYVQNSNIYVNSMYEFSFSFWIRI---NKWVSLPGYTI 117	
Db	882 PTKNQFTIFS---KPEVNIQNDYIYDKNKFSISFWIRPNDYDKIVLNINYEYTI 939	
QY	118 IDSVK-NNSGWSIGIISNLFVTLKQNEDESQSINFSYDINNAPGY-NKWFYFVITNNM 175	
Db	940 INCMRDNNSGWSKSLNHNHIIWTDQNRINQKLVFYGNANGISDYINKWFVITINDR 999	
QY	176 MGNWKIYINGKLIDTTKVVELGTFNFSKTIIFEINKIPDTGLITSDSDINMWRDVFIF 235	
Db	1000 LGDSKLYINGHLIDQKSIILNLGNIHVSNDLFIKIVNCSYTRYI-----GIRYENIF 1050	
QY	236 AKELDGKDINILFNSLQYTVNPKDYWGNDLRNKEYYVMNIDYLNRYM--YANSRQIVFN 293	
Db	1051 DKELDETEIOTLYSNEPNINILKDFWGNVLLYDKGYLLNVLKPNFIDRKDSTLSINN 1110	
QY	294 TRRN---NNDNFEGYKIIKTRR-GRNTDTRVVRGDIILYFDMTINNKAYNLFMKNETMYA 349	
Db	1111 IRSTILLANLKYSGIKVQIRVNDSTDRFYRKNQDQYVINYINSSSYSL-----YA 1163	
QY	350 DNHSTEDIYAIGREOT--KDINDNIIFOIQPMNNTYVYASQIFKSNFGNIGNISGICSG 407	
Db	1164 DTNITD-----KRTIIKSSSGRHFNGQVVMNSVGNCTMNFKN--NGNN-----IG 1209	
QY	408 TYRFR----LGGDWYRHNLVPTVKQGNYSALLESTSFH---WGWF 446	

1210 ILGFKADTVVASTWY-----YTHMRDHTNSNGCFWNPFI 1242

DB

RESULT 12

Q08077

ID C08077 PRELIMINARY; PRT: 1291 AA.

AC Q08077;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DT Botulinum neurotoxin type B (EC 3.4.24.-) (BONT/B).

GN BONT/B.

OS Clostridium botulinum.

OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OS Clostridiales; Clostridiaceae; Clostridium.

OX NCBI\_Taxid=1491;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EKLUND 17B ATCC25765;

RA MEDLINE=94122659; Pubmed=7764370;

RX Hutson R.A., Collins M.D., East A.K., Thompson D.E.;

RT "Nucleotide sequence of the gene coding for non-proteolytic

RT Clostridium botulinum type B neurotoxin: comparison with other

RT Clostridial neurotoxins.";

RL Curr. Microbiol. 28:101-110(1994).

CC -/- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER

CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD

CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT

CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC

CC ENDOPEPTIDASE THAT CLEAVES SYNAPTOSOMAL-2.

CC -/- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A

CC HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,

CC WHILE THE N- AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL

CC FORMATION AND TOXIN BINDING, RESPECTIVELY.

CC -/- SUBCELLULAR LOCATION: SECRETED.

CC -/- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF

CC BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

CC -/- SIMILARITY: HIGH WITH OTHER BOTULINUM NEUROTOXINS AND WITH TETANUS

CC NEUROTOXIN.

CC -/- SIMILARITY: TO OTHER ZINC METALLOPROTEINASES IN THE ACTIVE SITE

CC REGION.

DR EMBL; X71343; CAA50482.1; -

DR HSP; P10845; 3BTA.

DR MEROPS; M27.002; -.

DR InterPro; IPR000395; Bontoxilysin.

DR InterPro; IPR000130; Zn\_MTPeptidase.

DR Pfam; PF01742; Peptidase\_M27.1.

DR PRINTS; PR00760; Bontoxilysin.

DR PRODOM; PD001963; Bontoxilysin; 1.

DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.

KW Neurotoxin; Transmembrane; Hydrolyase; Metalloprotease; Zinc.

SQ SEQUENCE 1291 AA; 150513 MW; 71BCAF23D69FAA CRC64;

Query Match 24.8%; Score 593; DB 2; Length 1291;

Best Local Similarity 32.9%; Pred. No. 2.9e-23;

Matches 162; Conservative 87; Mismatches 159; Indels 84; Gaps

Qy 3 IPRNIFYTNNLLKDINEYFNFNINDSKILSNQKNKPLVDTGSYNAEVSEGDVQLNP 62

||||: |||: | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 836 I PFDLSYNTIEILIFNKYNSEIILNNILRLRYDRNNLIDLSGYAKVEYDGVKND 895

Qy 63 IFFPDFKLGGSGEDRGKVIYTONENIYNVSMYESISFWIRNKW---VSNL--PGYT 116

||||: |||: | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 896 --KNQFKLTGSAD--SKIRVTQNQNIIFNMLDFDSVFWIRPIRYNRDDIQNYTHNEYT 951

Qy 117 IIDSVRKNSGWSIGITSNFLVFTLKQNEQSEGSINFYSIDSNAPGY-NKWFVFTVTNNM 175

||||: |||: | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 952 IINCCKNSGKKV.SIRGNRIIWLIDINGTKTSVFVEYNREDISEYNINRPFVITNN- 1010

Qy 176 MGNKKVIYNGKLDITTKVKELTCINFESKITTEINKIPDTGLITSDSNIN-MWIRDFYI 234

||||: |||: | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 1011 LDNAKIYNTGLTESNMDIKDIGVYNGITFTKLD-----GVDRTQFIWKKYFSI 1061

```

Query Match      24.0%; Score 575; DB 2; Length 441;
Best Local Similarity 33.5%; Pred. No. 7.9e-33;
Matches 161; Conservative 78; Mismatches 159; Indels 82; Gaps 21;

Qy 10 YTNWLLKDIIEYFNFNNDKIIISLQNRKKTFLVDTSGYNAEVSEEGDVLQNPITPFPQK 69
    ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 1 YTNNTILIEIFNKYSKILNNLILNRYDRNNLIDLSGYCANVEYDGVLEND--KNQFK 58

Qy 70 LGSSGDEKVIYQTNQNIYVNSMYESISFSWIRINKW----VSNL--PGVTIIDSVKN 123
    || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 59 LTGS--TNSBIRVTQNIIFNSMFLDFSVGFWRIBPKYKNDGIQNIYHNEYTIINCIN 116

Qy 124 NSGWSIGIISNLFVLTKQNESEQISNFSDIENNAPGY--NKWFFVTYNNMGNMKIY 182
    ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 117 NSGWKISGRNRIIWLTDIDNGTKTSVPFEYSIREDISDYINRWFFVTINN--SDNAKIY 175

Qy 183 INGLDLDTYKVELGFINFSKTIPEINKIPEDTGLITSDSDNIN--MWIRDFYIFAKELDG 241
    ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 176 INKLESNIDIDKIGVIANGIILPKD-----GDIDRTQFIWMKYFSIFNLSUQ 226

Qy 242 KDNILFNLSQTYNNVKDYWGNDLREYKEYVVAIDYINRMY-----ANSRQIVFNREN 297
    ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 227 SNIKEIYKTSYSEYLFKDFGWNPLMYNKEYYMFNAGNKNYSYIKLKKDSSVGEIILTRSKYN 286
    ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|

```

[illegible]



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: November 7, 2002, 17:45:16 ; Search time 1950 Seconds  
(without alignments)  
3737.420 Million cell updates/sec  
Title: US-09-910-186a-10  
Perfect score: 2391  
Sequence: 1 MTIFNFSTNNSLLKDI.....NYASLLESTHGWGFPVSE 450

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2.1/USPT.spool/US09910186/runat\_04112002\_111617\_20848/app\_query.fasta\_1.647  
-DB=EST -QEXT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09910186\_cgn\_1\_17899\_drnat\_04112002\_111617\_20848 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_othr:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	127.5	5.3	870	17	AZ675343 ENTIG64TF
2	122.5	5.1	929	17	BH167076 ENTQX64TR
3	122	5.1	3707	14	C83838 C83838 Dict
4	119.5	5.0	917	17	AZ684830 ENTHZ58TF
5	119	5.0	649	9	AT297914 LP12358.5
6	117	4.9	573	17	AA549915 098m3 gm
7	116	4.9	562	13	BJ360994 BJ360994
8	116	4.8	641	12	BF297786 056PBC01
9	114.5	4.8	664	10	AW584247 N210227e
10	114	4.8	653	17	AA50238 135m3 gm
11	113.5	4.7	846	17	AZ689135 ENTGX61TF
12	112	4.7	549	13	BI813874 PfESToa1
13	112	4.7	943	17	BH134547 ENTOL08TF
14	110	4.6	500	9	AU087833 AU087833
15	110	4.6	561	13	BM275413 PfESTOa6
16	110	4.6	690	13	BM162769 EST565303
17	110	4.6	945	17	AL404242 T7 end of
18	109.5	4.6	696	13	BM162617 EST565140
19	109.5	4.6	720	13	BM169137 EST571660
20	109.5	4.6	781	13	BM166487 EST569010
21	109	4.6	571	13	BM274132 PfESTOa6
22	108.5	4.5	597	14	BQ597280 PfESTOa3
23	108	4.5	638	14	BM815053 EST593147
24	108	4.5	754	17	AA503316 1457m3 gm
25	108	4.5	796	13	BM159196 EST561719
26	108	4.5	877	17	AZ683138 ENTML73TR
27	107.5	4.5	767	13	BM162128 EST564851
28	107.5	4.5	920	17	AZ547074 ENTEN03TF
29	107	4.5	568	13	BM275819 PfESTOa8
30	107	4.5	610	14	BQ595977 PfESTOa2
31	107	4.5	722	13	BM164970 EST567493
32	107	4.5	1225	9	AB012854 AB012854
33	106.5	4.5	662	17	AZ524207 228PBD12
34	106	4.4	561	13	BM274447 PfESTOa4
35	106	4.4	732	13	BM159156 EST561679
36	106	4.4	950	17	AZ693368 ENTJUL19TR
37	105.5	4.4	688	12	BG604058 EST503148
38	105.5	4.4	694	12	BG604057 EST503147
39	105.5	4.4	698	13	BM163799 EST566322
40	105.5	4.4	764	13	BJ430749 BJ430749
41	105	4.4	1110	17	CNS06PWT
42	104.5	4.4	622	17	AZ526922 261PBC09
43	104.5	4.4	687	17	AZ526437 255PBE01
44	104.5	4.4	750	13	BM164094 EST566617
45	104.5	4.4	812	13	BM163086 EST565609

ALIGNMENTS

RESULT 1  
AZ675343  
LOCUS  
DEFINITION  
ENTIG64TF Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000  
ACCESSION  
AZ675343  
VERSION  
AZ675343.1 GI:11812609  
KEYWORDS  
GSS.  
SOURCE  
Entamoeba histolytica.  
ORGANISM  
Entamoeba histolytica.  
REFERENCE  
1 (bases 1 to 870)  
AUTHORS  
Loftus, B., Van Aken, S. and Fraser, C.  
TITLE  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library



JOURNAL  
COMMENT

Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Forward  
Class: shotgun  
High quality sequence start: 23  
High quality sequence stop: 724.

## FEATURES

source  
1. .870  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: PHOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G. and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:430.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 376 a 112 c 126 g 256 t  
ORIGIN

## Alignment Scores:

Pred. No.: 4.26e-05 Length: 870  
Score: 127.50 Matches: 71  
Percent Similarity: 37.58% Conservative: 44  
Best Local Similarity: 23.20% Mismatches: 116  
Query Match: 5.33% Indels: 75  
DB: 17 Gaps: 13

US-09-910-186a-10 (1-450) x A2675343 (1-870)

Qy 7 IlePheSerTyThrAsn-----AsnSerLeuLeuLys 17  
Db 18 ATCTTTACATTAACATCTATTTGCGATGATAACAAGATCAATATAACAACATTCACAGAA 77  
Qy 18 AspIleAsnGluTyRPhEAsnAsnIleAsnAspSerLysIleLeuSer----- 34  
Db 78 GACATAATAATAATACATTAACATAATGTTCCACCAATAAACAACACACAGTGCAC 137  
Qy 35 -----LeuGlnAsnArgLysAsnThrLeuValAspThrSerGlyTyRAsnAlaGluVal 52  
Db 138 CCACATAATAATGTAACATGATGTTTCCAAAATACCTTCCAAAATACACAGATGTG 197  
Qy 53 SerGluGluGlyAspValGlnLeuAsnProIlePhePhePheAspPheLysLeuGlySer 72  
Db 198 TCIGAA-----CCGAATGTCATTTGAAGGAAGTATAAGATTG 236  
Qy 73 SerGlyGluAspArgGlyLysValIleValThrGlnAsnGluAsnIleValTyRAsnSer 92  
Db 237 GAAAGTCATGAAATGAAATGAAATGAAATATATATATATATATATATATATATATATAT 296  
Qy 93 MetTyRGlusPheSerIleSerPheTTPile----- 103  
Db 297 AAGATGAGATGATGCATTTGTTATCTAGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 356  
Qy 104 ---ArgIleAsnLysTrpValSerAsnLeuProGlyTyThrIleIleAspSerValLys 122  
Db 357 ACATTATTGAATAGTTTGTGAATGATTATATATATATATATATATATATATATATATATAT 416

Qy 123 AsnAsnSerGlyTrpSerIleGlyIleIleSerAsnPheLeuValPheThrLeuLysGln 142  
Db 417 TAT-----ATAATTATAAT-----GAAGATAATTATGACAA 449  
Qy 143 AsnGluAspSerGluGlnSerIleAsnPheSerTyRAspIleSerAsnAlaProGly 162  
Db 450 CGTAAAGATCAATCTAAAGTCAACAAAGTCAA----- 482  
Qy 163 TyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIleTyr 182  
Db 483 -----GTAACAATTTATAAT-----ATTAAAGAACAAAAAGAACAA 518  
Qy 183 IleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSer 202  
Db 519 CCACCAATCAAAATATATACATACACCAAGATTTGGAGGAAATTCGTTATCAT 578  
Qy 203 LysThrIleThrPheGluIle-----AsnLysIleProAspThrGlyLeu 217  
Db 579 ATGGAATTTACTAAACAAATAAAGAAGCATTTGAAATATAAGTATTAGATTAAATGCT 638  
Qy 218 Ile-----ThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyRilePhe 235  
Db 639 ATTGTTTGTTCGCAACATCAAGTATGTCGAGGCTAACACTAAGTCAACACTATATTT 698  
Qy 236 AlaLys-----GluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGluTyr 253  
Db 699 GGAATATTATCACTCAACTATTTTGGAAAGATGTAAAGAAATT-----TAT 743  
Qy 254 ThrAsnValLysAspTyRTrpGlyAsnAspLeuArgTyRAsn-----Lys 269  
Db 744 TGAATGCTTACATTCGTGATGGTAAACACCAACAATTTATATGCACTACATCAAA 803  
Qy 270 GluTyRTrpMetValAsn 275  
Db 804 GATGATTTTTCAGTCCAT 821

## RESULT 2

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

BH167076 929 bp DNA linear GSS 24-SEP-2001  
ENTOX64TR Entamoeba histolytica Sheared DNA Entamoeba histolytica  
genomic, DNA sequence.  
BH167076  
BH167076.1 GI:15740514  
GSS.  
Entamoeba histolytica.  
Entamoeba histolytica.  
Eukaryota; Entamoebidae; Entamoeba.  
1 (bases 1 to 929)  
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.  
Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library (2001)  
Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: shotgun  
High quality sequence start: 17  
High quality sequence stop: 592.  
Location/Qualifiers  
1. .929  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="Vector: PHOS1; Site\_1: Bst I; Constructed at The

FEATURES  
source



```
Qy 109 ValSerAsnLeuProGlyTyrThrIleileAspSerVallyAsnAsnSerGlyTrpSer 128
Db 1880 CAAGATGACCTATTGGT-----AGTAATAGTAGTAATCA 1915
Qy 129 IleGlyIleIleSerAsnPhLeuValPheThrLeuLysGln----- 142
Db 1916 AAGATTATCAAAATCAACTCAATGTTTCTCTATCAAGTCGATATCTGGTGGTAAT 1975
Qy 143 ---AsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIleSerAsnAlapro 161
Db 1976 TGTAATAGTAGTAATCAAGATTAATCAAAATTAACACTCAATGTTTCTCTATCAAGT 2035
Qy 162 GlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMetGlyAsnMetLysIle 181
Db 2036 GCATAT-----ACTGGTGGTAATTTTAACTAATACCATATC 2071
Qy 182 TyrIleAsnGlyLysLeuIleAspThr---IleLysVallySgluLeuThrGlyIleAsn 200
Db 2072 TATGCCAAAATAAGATTAATGATACCGGGTGAATATCAACCTTTTCAATGGGCTCA 2131
Qy 201 PheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIleThrSer 220
Db 2132 GCAAACAATACCATTTCTATGTCGAAATAAAGTTAATGATACCGGTGTGAAATATCA 2191
Qy 221 -----AspSerAspAsnIle----- 225
Db 2192 ACCTTTTCAATGGGCTCAGCAGATGTGTTTATGTGTCGACAAATTTTAACTTTAT 2251
Qy 226 -----AsnMetTyrPheArgAspPheTyrIlePheAlaLysGluLeuAspGlyLys 242
Db 2252 CAATTAATAATAAGATGGTGTACTGGTTCATCATATTTTAAATTTTATGTGTCGAA 2311
Qy 243 AspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsn----- 255
Db 2312 CAA-----TTTTTAACCTTTATCAATTAATAATGAATGGTGTACTGTTCCAT 2362
Qy 256 ValVallyAsp-----TyrTrp----- 261
Db 2363 ATTTTAAAAATAAAATTTATTTACTAGTAAAAAGTGGTATCACTCAAAATGAATGCCCA 2422
Qy 262 GlyAsnAspLeuArgTyr-----AsnLysGlnTyrTyr----- 272
Db 2423 GTATGTGATCAAGATCTTTATTACAAATAAATTTATTACTTACTAGTAAAGTGTATC 2482
Qy 273 -----MetValAsnIleAspTyrLeuAsn 280
Db 2483 ACTCAATGAATGCCAGTACTGATCAAGATCTTTATTAAACAATGAATCTGCGAA 2542
Qy 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsn----- 297
Db 2543 TCATTTGAATTTGTAATGTAATGAAACAATTTATTATTATACACAAATGTAATAAT 2602
Qy 298 AsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArgGlyAsnThrAsn 317
Db 2603 AAGAAGAATCTGCGAATCAATTTGAATTTGTA-----ATAATGAA 2644
Qy 318 AspThrArgValArgGlyValAspIleLeuTyrPheAspMetThrIleAsnAsnLysAla 337
Db 2645 AAAACA-----ATTATTAT---ATTACAAATGTAATAAT 2680
Qy 338 TyrAsnLeuPheMetLysAsn-----GluThrMetTyrAlaAspAsnHisSerThr 354
Db 2681 AAGAATAATCGATGTGAAATTTAGTTTCATCAACACTTTAT---GATTCACTACTGTAT 2737
Qy 355 GluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIle 374
Db 2738 AAA-----TCATTAGGTGTGTGTAATGTGACCCCGATGTT---GAAATTTAGTT 2785
Qy 375 PheGlnIleGlnProMetAsnAsnThrTyrTyr----- 385
Db 2786 -----TCATCAACACTTTATGATTCATCTACTCTGATAATCATTAGGT 2827
```

```
Qy 386 -----TyrAlaSerGlnIlePheLysSerAsn 394
Db 2828 GTTTGTGAATGTGCACCAGGTTTCTCTGGAATTTATGTAATCAATGTTTAAATGTGACA 2887
Qy 395 PheAsnGlyGluAsn-----IleSerGlyIle----- 403
Db 2888 GTTCAATGGGAATATGTAATCAACATGTGTTAAAGTTTCTCTGGAATTTATGTAAT 2947
Qy 404 ---CysSerIleGlyThrTyrArgPheArgLeuGly 414
Db 2948 CAATGTGTTAATGTGTACAGTTCAATGGGAAATGGT 2983

RESULT 4
LOCUS      AZ684830                917 bp      DNA      linear      GSS 14-DEC-2000
DEFINITION      ENTH2581F Entamoeba histolytica Sheared DNA Entamoeba histolytica
                genomic, DNA sequence.
ACCESSION      AZ684830
VERSION        AZ684830.1  GI:11821976
KEYWORDS       GSS.
SOURCE         Entamoeba histolytica.
ORGANISM       Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE      1 (bases 1 to 917)
AUTHORS        Loftus,B., Van Aken,S. and Fraser,C.
TITLES         Determination of clone end sequences from Entamoeba histolytica
                HML:IMSS sheared DNA library
JOURNAL        Unpublished (2000)
COMMENT        Contact: Brendan J Loftus
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0208
                Fax: 301 838 3543
                Email: bilofus@tigr.org
                Clones are derived from the Entamoeba histolytica HML:IMSS sheared
                DNA library
                Seq primer: M13-Forward
                Class: shotgun
                High quality sequence start: 24
                High quality sequence stop: 871.
FEATURES       Location/Qualifiers
                1..917
                /organism="Entamoeba histolytica"
                /strain="HML:IMSS"
                /clone_lib="Entamoeba histolytica Sheared DNA"
                /note="Vector: PHOS1; Site_1: Bst I; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from broth cultures of E. histolytica
                using a method described by Clark and Diamond (Clark,
                C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                method for isolate identification. Exp. Parasitol.
                77:450.). The DNA was mechanically sheared to give a
                tight size distribution (~2 kb). The v + i method used for
                the library construction is described in detail in Smith,
                H.O. and Venter, J.C. (Making small insert libraries for
                whole genome shotgun sequencing projects. In Genome
                Sequencing: A Practical Approach, eds. M. Vaudin and B.
                Barrell, Oxford University Press, 1999)."
```

BASE COUNT 390 a 104 c 153 g 270 t

ORIGIN

Alignment Scores:

Pred. No.:	0.000489	Length:	917
Score:	119.50	Matches:	77
Percent Similarity:	34.59%	Conservative:	61
Best Local Similarity:	19.30%	Mismatches:	116
Query Match:	5.00%	Indels:	145
DB:	17	Gaps:	17

US-09-910-186A-10 (1-450) x AZ684830 (1-917)

QY	Db	799	ATTATAGTGAC	.....AAGCAGAGAGATGATTATTACACAAATGTTTGGAAATAGGT	852
QY	362	LeuargGlu	.....GlnThrLysAspIleAsnAspAsnIlelle	374	
Db	853	ATGAAAGAGTTTGTGACTTTTATACACAAATTAGAAATATTATCCGAGAAATCATTT	909		
RESULT	5				
AI297914					
LOCUS		649 bp	linear	EST 19-APR-2001	
DEFINITION		LP12358.5prime LP Drosophila melanogaster larval-early pupal pot2			
ACCESSION		Drosophila melanogaster cDNA clone LP12358		Sprieme, mRNA sequence.	
VERSION		AI297914			
KEYWORDS		AI297914.1 GI:3947321			
SOURCE		EST.			
ORGANISM		fruit fly.			
REFERENCE		Drosophila melanogaster			
AUTHORS		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
TITLE		1 (bases 1 to 649)			
JOURNAL		Harvey, D., Brokstein, P., Hong, L., Evans-Hoim, M., Su, C., Tsang, G., Lewis, S., and Rubin, G. M.			
COMMENT		BDGP/HMI Drosophila EST Project			
		Unpublished (2001)			
		Contact: Stapleton, M.			
		BDGP			
		Lawrence Berkeley National Lab			
		One Cyclotron Rd, Berkeley, CA 94720, USA			
		Fax: 510 486 6798			
		Email: <a href="http://www.fruitfly.org/EST">http://www.fruitfly.org/EST</a> , <a href="mailto:est@fruitfly.berkeley.edu">est@fruitfly.berkeley.edu</a>			
		Plate: 123 row: E column: 10			
		High quality sequence stop: 600.			
FEATURES		Location/Qualifiers			
		1..649			
		/organism="Drosophila melanogaster"			
		/db_xref="taxon:7227"			
		/clone="LP12358"			
		/clone_lib="LP Drosophila melanogaster larval-early pupal pot2"			
		/sex="male and female"			
		/dev_stage="larvae-pupae"			
		/lab_host="DH5-alpha"			
		/note="organ: whole body; Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."			
BASE COUNT		180 a 183 c 134 g 152 t			
ORIGIN					
Alignment Scores:					
Pred. No.:		0.000312	Length:	649	
Score:		119.00	Matches:	56	
Percent Similarity:		39.47%	Conservative:	34	
Best Local Similarity:		24.56%	Mismatches:	84	
Query Match:		4.98%	Indels:	54	
DB:		9	Gaps:	13	
US-09-910-186A-10 (11-649) x AI297914 (11-649)					
QY	138	PhethrLeuLysGlnAsnGluaspSerGluGlnSerIleAsnPheserTyraSpilesSer	157		
Db	36	TTCTCTGCTGAAGGAGAGCGATGGATCCGACTCATCTTTGAAATATACGTGCCCCATTCC	95		
QY	158	AsnAsnAlaProGlyTyraSnLysTrpPhePheValThrValThrAsnAsnMetMetGly	177		
Db	96	.....TACACTACAGTGAGACAAATACTTTTGAAC	128		
QY	178	AsnMetLysIleTyrrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThr	197		
Db	129	ACCACGCCAAGGTTTATT	167		
QY	198	GlyIleAsnPheserLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeu	217		

```

Db 168 ACGTTCATCACTCAACAAATAAATGATGTGTGCTCAATATTACGACAGACCGGATAT 227
Qy 218 -----IleThrSerAspSerAspAsnIleAsnMetTrp-----IleArgAspPheTyr 233
Db 228 TACGCTGCACCTACTACCGACGACAT-----TGCATGCCATCCACACCGCTCTG 278
Qy 234 IlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyr 253
Db 279 ATCAGACCCCAAC-----ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGlu 270
Qy 254 ThrAsn-----ValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGlu 270
Db 309 AACATCGCCGCCAGATCGTGATCTGTC--AACTGCGCCAGCGCTGGATGTT 365
Qy 271 TyrTyrMetValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIle 290
Db 366 ACCTACACACCTTACCTTGGATGTGATC--GAGTATTGACGACGAGACCACTACATC 422
Qy 291 ValPheAsnThrArgAsnAsnAspPheAsnGluGlyTyrLysIleIleLys 310
Db 423 CCTGGACATCGGCTTCAAC-----GGCTTCAACTGACTGATC 464
Qy 311 ArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAsp 330
Db 465 CGCCTC-----GGAAACGACACG-----GCCGATTCAATTACTATATC 503
Qy 331 MetThrIleAsnAsnLysAlaTyrAsnLeuPheMetTyrAsnGlnThrMetTyrAlaAsp 350
Db 504 CAGACCCCTGACCAACAGGCTACATCAGCTGGCTTCATGAA-----GCATCG 554
Qy 351 AsnHisSerThrGluAspIleTyr 358
Db 555 AATGATACAGCCTTGGACATTAC 578

RESULT 6
AA549915
LOCUS
DEFINITION
  0988m3 gmbpFHB3.1, G. Roman Reddy Plasmodium falciparum genomic
  clone 0988m, DNA sequence.
ACCESSION
  AA549915
VERSION
  AA549915.1 GI:2320167
KEYWORDS
  GSS.
SOURCE
  malaria parasite P. falciparum.
  Plasmodium falciparum
  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
  1 (bases 1 to 575)
AUTHORS
  Dame, J.B., Arnot, D.E., Bourke, P., Chakrabarti, D., Christodoulou, Z.,
  Coppel, R., Cowman, A., Craig, A., Fischer, K., Foster, J., Goodman, N.,
  Hinterberg, K., Holder, A.A., Holt, D., Kemp, D., Lanzer, M., Lim, A.,
  Newbold, C., Ravetch, J.V., Reddy, G.R., Rubio, J., Schuster, S.M., Su
  X.-Z., Thompson, J.K., Vital, F., Wellens, T.E. and Werner, E.M.
  Current status of the Plasmodium falciparum genome project
  Mol. Biochem. Parasitol. 79, 1-12 (1996)
  97001675
TITLE
  Contact: Dame JB
  Dept. of Pathobiology, College of Veterinary Medicine
  University of Florida
  2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
  Tel: 352 392 4700
  Fax: 352 392 9704
  Email: damej@mail.vetmed.ufl.edu
  Seq primer: T3
  Class: shotgun.
FEATURES
  source
    1..575
    /organism="Plasmodium falciparum"
    /db_xref="taxon:5833"
    /clone="0988m"
    /lab_host="E. coli XL1-Blue"
    /note="vector: pBluescript SK(+); Genomic DNA, from
    asynchronous blood stage parasites of the cloned Honduran
  
```

```

BASE COUNT      286 a      45 c      68 g      175 t      1 others
ORIGIN

Alignment Scores:      0.000455      Length:      575
Pred. No.:      117.00      Matches:      49
Score:      36.73%      Conservative:      23
Percent Similarity:      25.00%      Mismatches:      76
Best Local Similarity:      4.89%      Indels:      48
Query Match:      17      Gaps:      8
DB:

US-09-910-186a-10 (1-450) x AA549915 (1-575)

Qy 173 AsnAsnMetMetGlyAsnMetLysIleTyrIleAsnGly----- 185
Db 48 AATAATATTATGGAAGAAAGAAAGAAATATTTAAATGCTGTTCTAAAGAACAAAGAA 107
Qy 186 LysLeuIleAspThrIleLysValLysGluLeuThrGlyIleAsnPheSerLysThrIle 205
Db 108 AATTGCGATGATATATATCCAAAAATGAATATATGATGATAAACAATAAACAACA 167
Qy 206 ThrPheGluIleAsn---LysIle-----ProAspThr 215
Db 168 CAATATTTTAAATAATGAAATTTTATTACATTTTAAATATATATGCAAAACAGAT 227
Qy 216 GlyLeuIleThrSerAsp-----SerAspAsnIleAsnMetTrpIleArgAspPheTyr 233
Db 228 TCATTAGATATAGATGATTTTAAATGATGAAAAAATATGACCTTGTTCGAATGTTA 287
Qy 234 IlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyr 253
Db 288 ATTAATAGAAACAACTTGTGATAAAATGATTTGGAATATTTATTATATGTTACAAITA 347
Qy 254 ThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMet 273
Db 348 AAAAATAATGAAAGAGAAAGAAATGAAAAAGAAATATTAATCATATGAATATTTCAG 407
Qy 274 ValAsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsn 293
Db 408 GAAATATTATACCAT---GATAAAATAGTATATACAAATCAATACCAATATGCTTAAT 464
Qy 294 ThrArgArgAsnAsn---AsnAspPheAsnGluGlyTyrLysIleIleLysArgIle 312
Db 465 ATGAATTATACACTCCCAATAGGTACACGAGGAGTAATAATATT----- 509
Qy 313 ArgGlyAsnThrAsnAspThrArgValArgGlyAspIleLeuTyrPheAspMetThr 332
Db 510 -----AATACTAATCCT----- 521
Qy 333 IleAsnAsnLysAlaTyrAsnLeuPheMetLysAsnGlnThrMetTyr 348
Db 522 -----AATGTTGNTCAAAGGAATGAAACAACTAT 551

RESULT 7
BJ360994
LOCUS
DEFINITION
  BJ360994 Dictyostelium discoideum cDNA library, CF Dictyostelium
  discoideum cDNA clone ddc8p21 5', mRNA sequence.
ACCESSION
  BJ360994
VERSION
  BJ360994.1 GI:19260589
KEYWORDS
  EST.
  Dictyostelium discoideum.
  Dictyostelium discoideum
  Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
  1 (bases 1 to 562)
  
```

HB3 isolate cultured in vitro, was digested with mung bean nuclease in the presence of 30% formamide at 500C (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). The ends of the fragments were ligated using T4 DNA polymerase, and the fragments were ligated to EcoR V-cleaved and dephosphorylated pBluescript SK(+). Recombinant plasmids transformed E. coli XL1-Blue.

```

SOURCE      Plasmodium berghei.
ORGANISM    Plasmodium berghei.
REFERENCE   1 (bases 1 to 641)
AUTHORS     Carlton,J.M.-R. and Dame,J.B.
TITLE       The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL     Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT     Contact: Dame JB
            Department of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611-0880, USA
            Tel.: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: T3.

FEATURES             Location/Qualifiers
     source            1..641
                     /organism="Plasmodium berghei"
                     /strain="ANKA clone HP (gametocyte producer)"
                     /db_xref="taxon:5821"
                     /clone_lib="Pb cDNA #17, Tommaso Pace, Marta Ponzi, and
                     Clara Frontali"
                     /dev_stage="asynchronous blood stage"
                     /lab_host="Wistar rats"
                     /note="vector: pBluescript II vector DNA, excised from
                     Lambda ZAP II.; Site_1: EcoRI; Site_2: XhoI; Total RNA was
                     extracted from asynchronous blood stage forms of the
                     cloned ANKA isolate of P. berghei, grown in wistar rats to
                     30% parasitemia and 2-5% gametocytemia. Contaminating host
                     white cells had previously been removed and final host
                     cell contamination estimated to be approximately 5%.
                     PolyA+ RNA was extracted and reverse transcribed using an
                     oligo dT-XhoI primer (Lambda ZAP II cDNA cloning kit,
                     Stratagene). Second strand cDNA was made following the
                     manufacturer's protocol. EcoRI adaptors were ligated to
                     the cDNA, and fragments were ligated into EcoRI/XhoI
                     digested vector."

BASE COUNT      332 a      68 c      85 g      156 t
ORIGIN

Alignment Scores:
Pred. No.:      0.000988      Length:      641
Score:          115.00      Matches:      62
Percent Similarity: 35.23%      Conservative: 31
Best Local Similarity: 23.48%      Mismatches: 83
Query Match:      4.81%      Indels:      88
DB:              12      Gaps:      14

US-09-910-186A-10 (1-450) x BF297786 (1-641)
QY 139 ThrLeuLysGluAsnGluAspSerGluClnSerIleAsnPheSerTyrAspIleSerAsn 158
Db 59 ACTAATAAAAGAGAGGAAACCAAGCAACCGAGAGATTCAAC-----AAAAAC 109
QY 159 AsnAlaProGlyTyrAsnLysTrpPhePheValThrValThrAsnAsnMetMetGlyAsn 178
Db 110 AATGAAATAATACAGGAATAAGTGG-----GGTGGT 139
QY 179 MetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeuThrGly 198
Db 140 TTGAGA-----TTGAG----- 145
QY 199 IleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGlyLeuIle 218
Db 146 ---AATGAATCTCTATAAAATTTTAGATAATCAAAATAGCTAAATGACAAAT----- 193
QY 219 ThrSerAspSerAspAsnIle-----AsnMetTrpIleArgAspPheTyrIle 234
Db 194 ATAATAAAGAGATGAATCTTACAGATGAAATAAT----- 229
QY 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIle---LeuPheAsnSerLeuGlnTyr 253
Db 230 -----AGGAGATTTGGAT---AAAGATGGTAAATTAGGAATTTCCGACAAATATTTT 280

```

Qy 254 ThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMet 273  
 Db 281 AACACGAAATAAATAATGGC---AATAATACACACGAAATTAACAAATCTAATATAAT 337  
 Qy 274 ValAsnLeuAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsn 293  
 Db 338 AAAATAATGATATATTCAGAAATCAAAATATAATTTACAGAAATTTTAAATAATAC 397  
 Qy 294 ThrArgArgAsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArg 313  
 Db 398 TACAATAAAATAGAAACAATTTTAAATGAAATAC----- 433  
 Qy 314 GlyAsnThrAsnAspThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIle 333  
 Db 434 GGGATCAAAATACACAAATCGGATAGGA-----GATGAACGAAC 475  
 Qy 334 AsnAsnLysAlaTyrAsn---LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHis 352  
 Db 476 AATAATAATAGTATAACATATTGTCATGAAACAATCTAAGCAACTATGCAATAGT--- 532  
 Qy 353 SerThrGluAspIleTyrAlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsn 372  
 Db 533 ----- 547  
 Qy 373 IleIlePheGlnIleGlnPrometAsnAsnThrTyrTyrAlaSerGlnIlePheLys 392  
 Db 548 -----AATCCAAAATGGGAAATACACGATTAACAACAATTTGCTGAT 592  
 Qy 393 SerAsnPheAsn 396  
 Db 593 GGCAATTTCAAT 604

RESULT 9  
 LOCUS AW584247 664 bp mRNA linear EST 07-SEP-2000  
 DEFINITION N210227e MHAM Medicago truncatula/Glomus versiforme mixed EST  
 library cDNA clone MHAM-1F21, mRNA sequence.

ACCESSION AW584247  
 VERSION AW584247.1 GI:7261301  
 KEYWORDS Medicago truncatula/Glomus versiforme mixed EST library.  
 SOURCE Medicago truncatula/Glomus versiforme mixed EST library  
 ORGANISM Eukaryota; mixed EST libraries.  
 REFERENCE 1 (bases 1 to 664)  
 AUTHORS Harrison M.J., Liu J., Peng H., Gonzales M., Ellis L., Town C.D.,  
 Bowman C.L., Craven M.B., Hansen T.S., Holt I.E. and Fraser C.M.  
 ESTs from roots of Medicago truncatula after colonization with  
 Glomus versiforme  
 Unpublished (2000)  
 CONTACT Harrison M.J.  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73401  
 Tel.: 580-223-5810  
 Fax: 580-221-7380  
 Email: mjharrison@noble.org  
 Other name: MHAM-1c-Cl1; Date: 3/14/00; Updated to the Database of  
 Expressed Sequence Tags (dbEST) on 04/27/00; More information is  
 available at 'http://chrisry.tamu.edu/medicago'.  
 Seq primer: T3.  
 Location/Qualifiers  
 1..664  
 /organism="Medicago truncatula/Glomus versiforme mixed EST  
 library"  
 /cultivar="Medicago truncatula genotype A17"  
 /db\_xref="taxon:119092"  
 /clone="MHAM-1F21"  
 /clone\_lib="MHAM"  
 /tissue\_type="roots colonized with Glomus versiforme"  
 /dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The library was  
 made from a mixture of RNA from each of these stages."

## FEATURES

source  
 1..664  
 /organism="Medicago truncatula/Glomus versiforme mixed EST  
 library"  
 /cultivar="Medicago truncatula genotype A17"  
 /db\_xref="taxon:119092"  
 /clone="MHAM-1F21"  
 /clone\_lib="MHAM"  
 /tissue\_type="roots colonized with Glomus versiforme"  
 /dev\_stage="Roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The library was  
 made from a mixture of RNA from each of these stages."

/lab\_host="E. coli strain XL0LR"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; cDNA was prepared from polyA+ enriched RNA from  
 roots harvested at 10, 17, 22, 31 and 38 days  
 post-inoculation with Glomus versiforme. The cDNA was  
 directionally ligated into the Unizap XR vector from  
 Stratagene and packaged using Gigapack III Gold packaging  
 extracts. Plasmids containing cDNA inserts were excised  
 from the recombinant lambda-Zap phage using Ex-assist  
 helper phage and propagated in XL0LR cells."  
 BASE COUNT 297 a 142 c 96 g 129 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.00122 Length: 664  
 Score: 114.50 Matches: 73  
 Percent Similarity: 31.61% Conservatives: 37  
 Best Local Similarity: 20.98% Mismatches: 95  
 Query Match: 4.79% Indels: 143  
 DB: 10 Gaps: 19

US-09-910-186A-10 (1-450) x AW584247 (1-664)

Qy 4 ProPheAsnIlePhe---SerTyrThrAsnAsnSerLeuLysAspIle----- 19  
 Db 4 CCAGCCACACACTTACCAATCATATGAAAAACTTCGTACACGATCAATCAACAAC 63  
 Qy 20 ---IleAsnGluTyrPheAsnAsnIleAsnAspSerLysIleLeuSerLeuGlnAsnArg 38  
 Db 64 AAGTTAATAATTACTTCACTACCTACCAAGATTCGAC----- 102  
 Qy 39 LysAsnThrLeuValAspThrSerGlyTyrAsnAlaGluValSerGluGluGlyAspVal 58  
 Db 103 -----AACAGTACCAGTAAGTTAGTGACACA 132  
 Qy 59 GlnLeuAsnProIlePheProPheAspPheLysLeuGlySerSerGlyGluAspArgGly 78  
 Db 133 AAATACACACACA-----TCTTACAGCAGCAGCAACAAT----- 171  
 Qy 79 LysValIleValThrGlnAsnGluAsnIleValTyrAsnSerMetTyrGluSerPheSer 98  
 Db 172 -----AACTAGCAGACAACAAGTATCTTCTAGTCTGATGACTTAC 216  
 Qy 99 IleSerPheThrIleArgIleAsnLysTyrPheValSerAsnLeuPro-----GlyTyrThr 116  
 Db 217 -----ACAAA-----AAGTCCCAAGAAAGATATAAC 246  
 Qy 117 IleIleAspSerValLysAsnAsnSerGlyTyrPheSerIleGlyIleIleSerAsnPheLeu 136  
 Db 247 TCGAACGAAAAATCAGAACAAACAAC-----AACGATAACAAG 285  
 Qy 137 ValPheThrLeuLysGlnAsnGluAspSerGluGlnSerIleAsnPheSerTyrAspIle 156  
 Db 286 TATTCTTACAAACAACAACAGGATCTCTAGCACAACAAGTTTACAGAANA--- 342  
 Qy 157 SerAsnAsnAlaProGlyTyrAsnLysTyrPhePheValThrValThrAsnAsnMetMet 176  
 Db 343 -----GGATACAAC-----TCCATGGAACCCGGAACAAC 372  
 Qy 177 GlyAsnMetLysIleTyrIleAsnGlyLysLeuIleAspThrIleLysValLysGluLeu 196  
 Db 373 AACAAACAACAAGTATTCTTACAACAACAACAACAAGAT----- 411  
 Qy 197 ThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsnLysIleProAspThrGly 216  
 Db 412 -----TCTTCTAGCAACA-----AAGTTCCAGAAGAAGT 444  
 Qy 217 LeuIleThrSerAspSerAspAsnIleAsnMetIleArgAspPheThrIlePheAla 236  
 Db 444 ----- 444  
 Qy 237 LysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnVal 256  
 Db ----- 256

```

Db 445 -----TACAACTCAATGCAAAATCAGAAC--- 468
Qy 257 ValysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMetValAsnIle 276
Db 469 -----ACAACTAC----- 480
Qy 277 AspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArg--- 295
Db 481 -----GAGAAGTATAACTACAAACAC-----AAAGTTGCTGTTAATGACAAATAC 525
Qy 296 -----ArgAsnAsnAspPheAsnGluGlyTyrLysIleIleLysArgIleArg 313
Db 526 AGCTTCAAAAGTAAACAACTACAACTACAACTGAG-----AGG 564
Qy 314 GlyAsnThrAsnAspThrArgValArgGlyAspIleLeuTyrPheAspMetThrIle 333
Db 565 CAAGGAATGAGTCACACGAGAGTTATGGAAGTGGGAATAATTTTATGATGTTAACTCT 624
Qy 334 AsnAsnLysAlaTyrAsnLeuPhe 341
Db 625 GAGGAGAAATATATACCAACATTC 648

RESULT 10
AA550238/c 653 bp DNA linear GSS 05-DEC-2000
LOCUS 1358m3 gmbpfH3.1, G. Roman Reddy Plasmodium falciparum genomic
DEFINITION clone 1358m, DNA sequence.
ACCESSION AA550238
VERSION AA550238.1 GI:2320490
KEYWORDS GSS.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 653)
AUTHORS Dame,J.B., Arnot,D.E., Bourke,P., Chakrabarti,D., Christodoulou,Z.,
Coppel,R., Cowman,A., Craig,A., Fischer,K., Foster,J., Goodman,N.,
Hinterberg,K., Holder,A.A., Holt,D., Kemp,D., Lanzer,M., Lim,A.,
Newbold,C., Ravetch,J.V., Reddy,G.R., Rubio,J., Schuster,S.M., Su
,X.-Z., Thompson,J.K., Vital,F., Wellens,T.E. and Werner,E.
Current status of the Plasmodium falciparum genome project
Mol. Biochem. Parasitol. 79, 1-12 (1996)
97001675
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: T3
Class: shotgun.

FEATURES
source
1. .653
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone="1358m"
/lab_host="gmbpfH3.1, G. Roman Reddy"
/note="Vector: pBluescript SK(+); Genomic DNA, from
asynchronous blood stage parasites of the cloned Honduran
HB3 isolate cultured in vitro, was digested with HincII
nuclease in the presence of 30% formamide at 50oC (Vernick
, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic
Acids Research 16:5883-5896). The ends of the fragments
were ligated to EcoR V-cleaved and dephosphorylated
pBluescript SK(+). Recombinant plasmids transfected E.
Coli XL1-Blue."
193 a 84 c 50 g 326 t
BASE COUNT
ORIGIN
Alignment Scores: 0.00137 Length: 653
Pred. No.:

```

```

Score: 114.00 Matches: 49
Percent Similarity: 39.71% Conservative: 32
Best Local Similarity: 24.02% Mismatches: 81
Query Match: 4.77% Indels: 42
DB: 17 Gaps: 10

US-09-910-186a-10 (1-450) x AA550238 (1-653)
Qy 253 TyrThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyr 272
Db 645 TATAATGACAAATTAAG-----AAAAACAAAAAATAAGTATTATGAA 598
Qy 273 MetValAsnIleAsnTyrLeuAsnArgTyrMetTyrAla----- 285
Db 597 ATGATAAATAATACATATATGAATGAACATGGATATACAGATATTGAAATGACGTTTA 538
Qy 286 AsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGluGlyTyr 305
Db 537 AATAAAAAAATAACGATTGAATGATAGAGGCTCGTACCACATACATTGGATCAT----- 484
Qy 306 LysIleIleLysArgIleArgGlyAsnThrAsnAspThr----- 319
Db 483 -----ATTATTGTAAGTCATCATGGAATAGTTATGATAAATAATATACTAGTAAACAT 430
Qy 320 ArgValArgGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsn 339
Db 429 AATAGAGAAAGAAATCATATATAATGAGATGAAAGAAACACAGAAATAATA----- 379
Qy 340 LeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAla 359
Db 378 -----AAAAAAACACGTTATTGTAGATGGA-----AAAGATATGGAAGGA 337
Qy 360 IleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnPro 379
Db 336 ATAGGAAAGAAAGAAAGAAATAAATAATATGAAATATATATTTTAT----- 289
Qy 380 MetAsnAsnThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGlu--- 398
Db 288 ---ATAACTCATATAGTATATAATAATAGTCTCTATAGTAATAATAATAATGATATA 232
Qy 399 ---AsIleSerGlyIleCysSerIleGlyThrTyrArgPheArgLeuGlyAspTyr 417
Db 231 TATAGTGTGATTAATGATCATCTGTTAATAATACAAATATGATATCAGGAGTCCCAAGC 172
Qy 418 TyrArgHis-----AsnTyrLeuValProThrValLysGlnGly-----AsnTyr 432
Db 171 TATGCTCATCTACTTATTATAAACAAGTAAATGAATATATATCAAGGCTACCTAATAT 112
Qy 433 AlaSerLeuLeu 436
Db 111 AATAACATGATG 100

RESULT 11
AA550238/c 846 bp DNA linear GSS 14-DEC-2000
LOCUS ENTGX611F Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ689135
VERSION AZ689135
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 846)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:FMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208

```





```

source
1..549
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/clone_lib="DH10B (GeneHog, invitrogen, Inc.)"
/lab_host="DH10B (GeneHog, invitrogen, Inc.)"
/notes="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2: XhoI; Library was constructed by Deopam Chakrabarti. Total RNA samples were isolated from mixed stage zaponin(0.1%) lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of l Zapir vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper plasmid (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."
BASE COUNT      278 a      51 c      54 g      166 t
ORIGIN
1..549
0.00182      Length:      549
Pred. No.:      Matches:      46
Score:      112.00
Percent Similarity:      41.01%
Best Local Similarity:      25.84%
Query Match:      4.68%
DB:      13

US-09-910-186A-10 (1-450) x BI813874 (1-549)
QY 222 SerAspAsnIleAsnMetTrpIleArgSpPheTyrIlePheAlaLys----- 237
Db 60 AGTCAAAATATTCAAATATATATATGATGATAAAAGAAATATGAAAAAATAATCAACTA 119
QY 238 GluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThrAsnValVal 257
Db 120 AAAATGTCAGGAAAGATATATGATATATATCATCCAAATTTGGATAAATTAATGATAAA 179
QY 258 LysAspTyrTrpGlyAsnAspLeuArg----- 266
Db 180 AAAATTTTGTGGCAATGATATGTCACACTCAAAATAGTCCGAAAAATTTCTCTCTTT 239
QY 267 -----TyrAsnLysGlnTyrTyrMetValAsnIleAspTyrLeuAsn 280
Db 240 GAAAAACCAATGGAATAAATATTCAAACTATATCTTGTAT---AAATATTAAAT 296
QY 281 ArgTyrMetTyrAlaAsnSerArgGlnIleValPheAsnThrArgAsnAsnAsn 300
Db 297 CATGAAAGAGATCAAAACATAAATAGTGTACATAATTTAAATGATAATAATAATAAT 356
QY 301 PheAsnGluGlyTyrLysIleIleLysArgGlyAsnThrAsnAspThrArg 320
Db 357 AATAATAAGACAAATAATATATCTAAGATATAATAAATCATATAATATATACAA--- 413
QY 321 ValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyrAsnLeu 340
Db 414 -----GATATATATTAT-----AACATAAATAATTCATACAAATAT 449
QY 341 Phe-----MetLysAsnGlnThrMetTyrAlaAspAsnHisSerThrGluAspIleTyr 358
Db 450 TTGTGTAATATGCAAAAGAAATGGAATATGCTAAAAACAATAAT---GAGCATTTA--- 503
QY 359 AlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGln 376
Db 504 -----AAAAAAACTACAGAACAGATACATATACAA 536

RESULT 13
BH134547/c      943 bp      DNA      linear      GSS 07-AUG-2001
LOCUS

```

```

DEFINITION
ENTOI08TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION
BH134547
VERSION
BH134547.1 GI:15093608
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica.
ORGANISM
Entamoeba histolytica; Entamoebidae; Entamoeba.
REFERENCE
1 (bases 1 to 943)
Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
JOURNAL
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
COMMENT
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 33
High quality sequence stop: 719.
FEATURES
Location/Qualifiers
1..943
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pHOSt; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barelil, Oxford University Press, 1999)."
BASE COUNT      317 a      142 c      119 g      365 t
ORIGIN
Alignment Scores:
Pred. No.:      0.00464      Length:      943
Score:      112.00      Matches:      56
Percent Similarity:      37.14%      Conservative:      35
Best Local Similarity:      22.86%      Mismatches:      80
Query Match:      4.68%      Indels:      74
DB:      17      Gaps:      13

US-09-910-186A-10 (1-450) x BH134547 (1-943)
QY 89 ValTyrAsnSerMetTyrGluSerPheSerIleSerPheTrpIleArgIleAsnLysTrp 108
Db 771 ATCTATTCAGTAGATACACACATTTTCATTACCAATGG---AGTATTACAACATGG 715
QY 109 ValSerAsnLeuProGlyTyrThrIleLeaspSerValLysAsnSerGlyTrpSer 128
Db 714 GTA---AATTTACCA----- 703
QY 129 IleGlyIleIleSerAsnPheLeuValPheThrLeuLysGlnAsn---GluAspSerGlu 147
Db 702 ---CCAGTTGTTAGTCAATTATCTTTGTTCAAAATTTATCAAAATATAAATGAGCTGAA 646
QY 148 GlnSer-----IleAsnPheSerTyr 154
Db 645 CAAAGTTTTTTTGAATTAATATGAACCAAGTGGTTTAAGAGTTATATATGTTTCAAAA 586

```

```

Qy 155 AspileSerAsnAlaProGlyIleAsnLysTrpPhePheValThrValThrAsnAsn 174
Db 595 GTAATATTGATATATGCGCATACCAAGAAATGGAATCATATTGTT----- 538
Qy 175 MetMetGlyAsnMetLys-----IleTyrIleAsnGlyLysLeuIleAspThrIle 191
Db 537 TTAGTAGAAAGAAAGAAATACATATATATATATATATATATATATATATATATAT 478
Qy 192 LysValLysGluLeuThrGlyIleAsnPheSerLysThr----- 204
Db 477 AATAA-----ATAAATTATACCTCAACTCCATATCGTTGTTATTGGA 433
Qy 205 -----IleThrPheLysLeuIleAsnLysLysProAsp 214
Db 432 CATTCAAAAGAAATAAGTAATACCAATGCAGTTTAAAAATAGGACCATGAAAATG 373
Qy 215 ThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArgAspPheTyrIle 234
Db 372 TTGGAATATCCATTTCGTATCAACA-----ATTAGAGAGTTATATTTA 328
Qy 235 PheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyrThr 254
Db 327 TTAGGAGAAATGATCAAGGAAGT-----TTCCAAACAATTTGATTATTC 283
Qy 255 -----AsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrThr 272
Db 282 AGATATGAAGTTTTCACAAATTATGATCAAAATGACATCAATTAATCAAGTTTACT 223
Qy 273 MetVal---AsnIleAspTyrLeuAsnArgTyrMetTyrAlaAsnSerArgGlnIleVal 291
Db 222 AATATAAAATCACTTAATCAACATTTAATTAATTAATTAATTAATTAATTAATTA 163
Qy 292 PheAsnThrArgArg 296
Db 162 ATTAATACAGATCGT 148

RESULT 14
LOCUS AU087833 500 bp mRNA linear EST 27-JAN-2001
DEFINITION AU087833 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA
clone XPn5946, mRNA sequence.
ACCESSION AU087833
VERSION AU087833.1 GI:12389974
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7.
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 500)
AUTHORS Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
TITLE FULL-malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum
JOURNAL Nucleic Acids Res. 29 (1), 70-71 (2001)
MEDLINE 20574754
COMMENT Contact: Junichi Watanabe
Institute of Medical Science
The University of Tokyo, Department of Parasitology
4-6-1, Shirokanebaidai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5378
Fax: 81-3-5449-5410
Email: jwatanab@manage.ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suvama,A. and Sugano
,S. Construction and characterization of a full length-enriched and
a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES
source
location/Qualifiers
1..500
/organism="Plasmodium falciparum 3D7"
/strain="3D7"
/db_xref="taxon:36329"
/clone="XPn5946"
/clone_lib="Sugano Malaria cDNA library"
/dev_stage="erythrocytic stage"
BASE COUNT 243 a 35 c 74 g 145 t 3 others

```

## ORIGIN

```

Alignment Scores: 0.00279 Length: 500
Pred. No.: 110.00 Matches: 44
Score: 39.61% Conservative: 17
Percent Similarity: 28.57% Mismatches: 59
Best Local Similarity: 4.60% Indels: 34
Query Match: 9 Gaps: 5
DB:

```

US-09-910-186a-10 (1-450) x AU087833 (1-500)

```

Qy 234 IlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPheAsnSerLeuGlnTyr 253
Db 94 ATATTTCCTTTTGACACATCAGCTGAGAAATAATAATATTATACAAACGATAGTATATA 153
Qy 254 ThrAsnValValLysAspTyrTrpGlyAsnAspLeuArgTyrAsnLysGluTyrTyrMet 273
Db 154 AATAAT-----GGAATAGG----- 168
Qy 274 ValAsnIleAspTyrLeuAsnArgTyrMetTyrAla-----AsnSer 287
Db 169 ---AATGTAGATGTAGGAAATGAAAGATGTTTAGTTATAATATGAAGCATATAATAA 225
Qy 288 ArgGlnIleValPheAsnThrArgArgAsnAsnAspPheAsnGlu----- 303
Db 226 AGCATTTGTTTAAAAAATGATACATATATATAATGATTTTAAATAAAGATATATTAT 285
Qy 304 -----GlyTyrLysIleIleIleLysArgIleArgGlyAsnThrAsnAsp 318
Db 286 GACAAACATAATAGTAATGAAATAATATATGAAAGTAAATAATCAATATAATAATGAT 345
Qy 319 ThrArgValArgGlyGlyAspIleLeuTyrPheAspMetThrIleAsnAsnLysAlaTyr 338
Db 346 AATAAAATATTTATGTAGATAAAATTTCTTGATGAAAAATAATTTAATAAAGATAGA 405
Qy 339 AsnLeuPheMetLysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyr 358
Db 406 TCCAAATTTAGAAAGATTAGACACATTC-----AACGAGAGTGACAAGATTATAAT 456
Qy 359 AlaIleGlyLeuArgGluGlnThrLysAspIleAsnAspAsn 372
Db 457 GCATGTGCATATATGAAATAAATGATATATAATGATATAAT 498

```

```

RESULT 15
LOCUS BM275413 561 bp mRNA linear EST 20-DEC-2001
DEFINITION PFEStoa68e04.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.
ACCESSION BM275413
VERSION BM275413.1 GI:17968741
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7.
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 561)
AUTHORS Tang,K., Cole,K., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.

```

```

TITLE WashU Plasmodium EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library was constructed by R. Haywood. DNA sequencing by:
Washington University Genome Sequencing Center For information on

```

obtaining a clone please contact: L. David Sibley  
(sibley@borcim.wustl.edu), Washington University  
Seq primer: 40UP from Gibco  
High quality sequence stop: 424.

# FEATURES

## source

Location/Qualifiers  
1..561  
/organism="Plasmodium falciparum 3D7"  
/db\_xref="taxon:36329"  
/clone\_lib="Plasmodium falciparum 3D7 gametocyte cDNA library"  
/dev\_stage="gametocyte (stage III-V)"  
/lab\_host="DH10B (GeneHog, Invitrogen, Inc.)"  
/note="Vector: pBluescript SK plus; Site\_1: EcoRI; Site\_2: XhoI; The library was constructed by R Haywood. cDNAs were synthesized from gametocyte poly(A)+ RNA by oligo d(T) priming, size-selected and directionally cloned into the EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR lambda vector (Stratagene). The primary library was mass excised as phagemid using the ExAssist helper phage (Stratagene). Clones were mass excised using the ExAssist helper phage (Stratagene), the phagemids were precipitated with PEG 8000 and extracted with phenol/chloroform. Phagemid DNA was electroporated into DH10B cells. Clone availability: David Sibley, Washington University."

BASE COUNT 270 a 32 c 62 g 197 t  
ORIGIN

## Alignment Scores:

Pred. No.: 0.00341 Length: 561  
Score: 110.00 Matches: 56  
Percent Similarity: 39.67% Conservative: 40  
Best Local Similarity: 23.14% Mismatches: 74  
Query Match: 4.60% Indels: 72  
DB: 13 Gaps: 13

US-09-910-186a-10 (1-450) x BM275413 (1-561)

QY 176 MetGlyAsnMetLysIleThrIleAsnGlyLysLeu-----IleAspThr 190  
Db 4 TTAAGCAATATATTTATATAAATAATAAATCAAGTTAACTCCCATACCTAGAAAT 63  
QY 191 IleLysValIleGluLeuThrGlyIleAsnPheSerLysThrIleThrPheGluIleAsn 210  
Db 64 GTGTCATAAAGAACATTTTGTGTAACTACTAGTATTAAATATA----- 108  
QY 211 LysIleProAspThrGlyLeuIleThrSerAspSerAspAsnIleAsnMetTrpIleArg 230  
Db 109 -----TTATTACATTATGATGAAGAAATATAATAAT----- 141  
QY 231 AspPheThrIlePheAlaLysGluLeuAspGlyLysAspIleAsnIleLeuPhe----- 248  
Db 142 -----AACTAATAAAATAGATTAATGAAGTTACATAAACATATCATATGGTAAT 195  
QY 249 AsnSerLeuGlnIleThrAsnValIleLysAspTyrTrpGlyAsnAspLeuArgTyrAsn 268  
Db 196 AATAACAATAATATAATAATATATGTGTTATTATTAAACAGAGATTATTATTGTTAT 255  
QY 269 LysGluTyrIleMetValAsnIleAsp-----TyrLeuAsnArgTyrMetTyrAla 285  
Db 256 GAATTTTATGCAATTAATCATATGATGATGATGATGATGATGATGATGATGATGAT 315  
QY 286 AsnSerArgGlnIleValPheAsnThrArgArgAsnAsnAsnAspPheAsn---GluGly 304  
Db 316 AATAAAT-----ATTGCTAATAAATAAATAAATAAATAAATAAATAAATAAATA 366  
QY 305 TyrLysIleIleLysArgIleArgGlyAsnThrAsnAspThrArgValArgGlyGly 324  
Db 367 AATAAACCTTTTACAAATATGATTAACAAATGAGAT----- 405  
QY 325 AspIleLeuTyrPheAspMetThr---IleAsn---AsnLysAlaTyrAsnLeuPheMet 342  
Db 406 ---GTATGGCAATATGATATTTCAAAAAATAAATAAATAAATAAATAAATAAATAAATA 462

QY 343 LysAsnGluThrMetTyrAlaAspAsnHisSerThrGluAspIleTyrAlaIleGlyLeu 362  
Db 463 AAAAGAGGAAATATG----- 477  
QY 363 ArgGluGlnThrLysAspIleAsnAspAsnIleIlePheGlnIleGlnProMetAsnAsn 382  
Db 478 -----AATACAAAAATATA-----AAT 495  
QY 383 ThrTyrTyrTyrAlaSerGlnIlePheLysSerAsnPheAsnGlyGluAsnIleSerGly 402  
Db 496 ACATATATATATGGAATAATATATTT-----TTTGAATGAATAATATATATATGAT 546  
QY 403 IleCys 404  
Db 547 ATATGT 552

Search completed: November 7, 2002, 19:10:18  
Job time : 1959 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 11:55:50 ; Search time 2773 seconds  
(without alignments)  
14388.729 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371

Sequence: 1 gaattcacgatgaccatccc.....tctccgagtaataaggaaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vi.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rtd.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgt\_hum.\*  
40: em\_hgt\_mus.\*  
41: em\_hgt\_other.\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	665.2	48.5	4479	7	CBCPHGC1	X53751 Clostridium
2	665.2	48.5	4592	7	CSTCITOX	D90210 Bacterioph
3	665.2	48.5	4712	1	CBCPHNC1M	X71126 C.botulinu
4	665.2	48.5	9613	1	CBCPTOX	X62389 Botulinum b
5	665.2	48.5	9689	1	CBP1NCTC1	X66433 Clostridium
6	665.2	48.5	11747	1	AB061780	AB061780 Clostridi
7	665.2	48.5	12297	7	CBCBONT	X72793 Clostridium
8	449.8	32.8	11631	1	AB037920	AB037920 Clostridi
9	448.2	32.7	4049	1	CLOTNGCO	D38442 Clostridium
10	189	13.8	151705	2	AC117900	AC117900 Rattus no
11	188.4	13.7	176351	2	AC117007	AC117007 Rattus no
12	183	13.3	169163	2	AC115666	AC115666 Rattus no
13	180.2	13.1	163034	2	AC099432	AC099432 Rattus no
14	170.2	12.4	165337	2	AC096212	AC096212 Rattus no
15	169.4	12.4	1402	6	AR000030	AR000030 Sequence
16	169.4	12.4	1402	6	AR169141	AR169141 Sequence
17	169.4	12.4	1402	6	AX036246	AX036246 Sequence
18	168.8	12.3	1330	6	AR000029	AR000029 Sequence
19	168.8	12.3	1330	6	AR169140	AR169140 Sequence
20	168.8	12.3	1330	6	AX036243	AX036243 Sequence
21	168.8	12.3	1338	12	XXU22962	U22962 Synthetic b
22	168.8	12.3	97683	2	AC116548	AC116548 Dictyoste
23	165.6	12.1	236542	2	AC096320	AC096320 Rattus no
24	164.6	12.0	256774	2	AC116964	AC116964 Dictyoste
25	160.2	11.7	149228	2	AC113774	AC113774 Rattus no
26	158.8	11.6	115758	9	AC104634	AC104634 Homo sapi
27	157	11.5	108476	2	AC094507	AC094507 Rattus no
28	156.8	11.4	158063	9	AP001046	AP001046 Homo sapi
29	156.8	11.4	340000	9	AP001751	AP001751 Homo sapi
30	155.4	11.3	173540	2	AC119697	AC119697 Rattus no
31	155.2	11.3	131346	2	AC119558	AC119558 Rattus no
32	154.6	11.3	169966	2	AC126213	AC126213 Rattus no
33	153.2	11.2	1299	12	AF251281	AF251281 Synthetic
34	153	11.2	7921	3	AF153362	AF153362 Dictyoste
35	151.6	11.1	1313	6	A58946	A58946 Sequence 6
36	151.6	11.1	39369	2	AC115681	AC115681 Dictyoste
37	150.6	11.0	186278	9	AC079176	AC079176 Homo sapi
38	147.6	10.8	229502	2	AC126871	AC126871 Rattus no
39	145.8	10.6	108177	2	AC096493	AC096493 Rattus no
40	145.8	10.6	125026	10	AC091616	AC091616 Rattus no
41	144.2	10.5	156533	2	AC117070	AC117070 Dictyoste
42	143.8	10.5	157141	3	AC016445	AC016445 Drosophil
43	136.6	10.0	147671	2	AC105625	AC105625 Rattus no
44	136	9.9	5163	6	AR097042	AR097042 Sequence
45	136	9.9	5318	6	AR097041	AR097041 Sequence

# ALIGNMENTS

RESULT 1  
LOCUS CBCPHGC1  
DEFINITION Clostridium botulinum C phage gene for C1 neurotoxin.  
ACCESSION X53751  
VERSION X53751.1 GI:14905  
KEYWORDS C1/neurotoxin; neurotoxin; toxin.  
SOURCE Clostridium botulinum C phage.  
ORGANISM Clostridium botulinum C phage  
Viruses.  
REFERENCE 1 (bases 1 to 4479)  
AUTHORS Popoff M.R.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-1990) Popoff M.R., Institut Pasteur, Anaerobies,  
25 rue du Docteur Roux, 75724 Paris Cedex 15, France

Pred. No. is the number of results predicted by chance to have a

REMARK revised by [3]  
REFERENCE 2 (bases 1 to 4479)  
AUTHORS Hauser, D., Eklund, M.W., Kurazono, H., Binz, T., Niemann, H., Gill, D.M., Boquet, P. and Popoff, M.R.  
TITLE Nucleotide sequence of Clostridium botulinum C1 neurotoxin  
JOURNAL Nucleic Acids Res. 18 (16), 4924 (1990)  
MEDLINE 90370487  
PUBMED 2204031  
REMARK revised by [3]  
REFERENCE 3 (bases 1 to 4479)  
AUTHORS Popoff, M.R.  
TITLE Direct Submission  
JOURNAL Submitted (24-SEP-1990)  
FEATURES Location/Qualifiers  
source 1..4479  
/organism="Clostridium botulinum C phase"  
/strain="phage from 468C botulinum C. strain"  
/db\_xref="taxon:12448"  
198..203  
/notes="pot. ribosome binding site"  
214..4089  
/notes="C1 neurotoxin (AA 1-1291)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA37780.1"  
/db\_xref="GI:14906"  
/db\_xref="SWISS-PROT:P18640"  
/translation="VFITINNYSDVDNKNILYLDTHLNTLANEPEKAFRTGNIW  
VDFRSRNPMLNPKPPTVSPKSGYDDPNVLSDSKDFELKEIKLKRINSREI  
GEEILYSLSDIPFNQNTINTDFDVFNSVDVKTQGNWVYKGSINPVSIIIG  
PRENIIDPSTFKLNTNFAAGGFGALSIISPRFMTYSNATNDVGGREFSKE  
FCMDPILILHNLHAMHLYGAIAPNQTSSVTSNIFYSQYNVLEYIAEYAFGP  
TDILIPKARKYFEKALDYRSIAKRLNITFANPFSFKYIGYKQKLKRYKRVV  
ESGEVTVNRKFNVELYELTQFTFENYAKIYNVONRKIYLSNVYPTVANIILDDV  
VDIQNGFNPKSLNVLFGONLSRNPALRKVNPEMLYLFKPCRAIDGRSLYNT  
LDCELLYKNTDLPFIDGLSDVKTDIFLRKIDNESTEVIYYPDNVSDVILSKTSE  
HGQDLNYSILSEILPGENQVFDNRTQNDVILNYYLSEKLSDNVEDFTFR  
SIEALDNKAVYTFPTLANKVAGVGGGLFWANDVDFEFTNILRKDILDKISD  
VSAIPIYGPALNISVSRGTFEAFVGTIILEAPFEFTIPALGAFYVSKVOE  
RNEIITKIDNCEKLRKWSYEMWGTWLSRIITQFNNTSYQMDSLNVOAGAIA  
KIDLEYKYSKGSKENIKSQVENLKNLSLDKISEAMNINIKFIRECVTYIFKMLPK  
VIDELNFRNTKAKILINDSHNIIIVGEVDKLAKNVNSFQNTIPENIFSYNLSL  
LKDIINEYNINDKISLQNKNTLVDTSCYNAEVSEGDVQLNFIKPFDFKLGSS  
GERGKAVIVQENIYNVNSWYSFISFWIRINKVSNLPQYTIIDSVKNSGSIIGI  
ISNPLFTLKQNSDSQSINFSDISNAPGPKWFFVTVNNMGMNKITKINGKILID  
TIKVLGTGINSCTITFEIKIPDTGLTSDSNINMIRDFYIFAKELDGDKNIL  
FNSLQYTNVYKWDYNDLRYNKEYVNIIDYLNRYNANSRQIVENTERNNDNFEGY  
KIIIKRIRGNTDVRGGDILYFDMTINKAYNLFMKNETMYADNHSSTEDIIAIGLR  
EQKIDINDIIFIOIPMNNTYVYASQIFKFSNNGENISGICSIGHYRFRIGDGNVHN  
YLVPYKQGNVSLLESTHMGFVYVSE"  
BASE COUNT 1817 a 455 c 693 g 1514 t  
ORIGIN  
Query Match 48.5%; Score 665.2; DB 7; Length 4479;  
Best Local Similarity 68.3%; Pred. No. 1.6e-92;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
QY 13 ACCATCCATTCAACACTCTCTCCACCAACAACTCCCTGTGAGGACATCAAC 72  
DB 2740 ACAATACCCCTTTAATTTTTCATACATAATAAATCTTTTATTAAGAATATAATTAAT 2799  
QY 73 GAGTACTTCAACATCAACGACTCCCAAGATCTGCTCCCTCGAGAACCGTAAGACACC 132  
DB 2800 GAATATTTCAATATATTAATGATTCAAAAATTTTGAGCCCTACAAACAGAAAAATCT 2859  
QY 133 TTGTCGACACTCCGGTTACAAACCGGAGGTCTCCGAGGAGGTTGACGTCGAGCTGAAC 192  
DB 2860 TTAGTGGATACATCAGGATATAATGCAAGAGTCTGAGTGAAGAAGCGATGTTTCAGCTTAAT 2919  
QY 193 CCAATCTCCATTCGACTTCAAGCTGGTTCCTCCGTCGAGGACGAGGTGAAGGTCACT 252  
DB 2920 CCAATATTTCCATTGCTTTAAATAGTAGTTTCAGGGGAGGATAGAGGTAAAGTTATA 2979



QY	973	AGAGGTGGTGACATCCTGTAATTCCGACATGACTACTCAACAACAAAGGCCTCAACCGTGTTC	1032
DB	3807	CGAGGAGGAGATATTTTATATTTTGGATGACAATTAATAAGAACAAAGCATATAATTTGTTT	3866
QY	1033	ATGAAGAACGAGACCATGTACGCCGCGAACAACCACTCCACCGAGGAGACATCTAGCCCATCGGT	1092
DB	3867	ATGAAGAATGAAACTATGTATGCAGATAATCATAGTACTGAAGATATATATGCTATAGGT	3926
QY	1093	CTGCGGTGACGACCAAGGACATCAACGACAAACATCATCTTCCAGATCCAGCCCAATGAAC	1152
DB	3927	TTAAGAGAACAAACAAGGATATAAATGATAATATATTATTTCAAATACAACCAATGAAT	3986
QY	1153	AACACATTACTACTACGCTTCCCAGATCTTCCAAGTCCCAACTCTCAACGGTGAGAACATCTCC	1212
DB	3987	AATACTATTATTACGCATCTCAATATTTAATCAAAATTTAATGGAGAAAATATTTCT	4046
QY	1213	GGTATCTGTTCCATCGGPACCTCAGATATCCGCTCTGGGTGGTGACTGGTACGACACAAAC	1272
DB	4047	GGAATAIGTTCAATAGTAGTACTATGCTGTTTTAGACTTTGGAGGTGATGGTATAGCACAAAT	4106
QY	1273	TACTTGGTTCCAACTGTCAACGAGGGTAACTACGCTCTTGGTGGAGTCCCACTTCCACC	1332
DB	4107	TATTTGGTCCCTACTGTGAACGAGGAAATATGCTTCATTATTAGAACATCAACATCAACT	4166
QY	1333	CAGTGGGATTCTGCCAGTCTCCGAGTAA	1362
DB	4167	CATTGGGGTTTGTACCTGTAAGTGAATAA	4196

RESULT 3  
CBPNCNClM

LOCUS C.botulinum phage C mutant gene for neurotoxin type Ci.  
DEFINITION X71126 BCT 23-DEC-1993  
ACCESSION X71126 linear  
VERSION X71126.1 GI:414619  
KEYWORDS mutant; neurotoxin.  
SOURCE Clostridium botulinum.  
ORGANISM Clostridia; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

REFERENCE .  
AUTHORS Hauser,D.  
TITLE Direct Submission  
JOURNAL Submitted (03-MAR-1993) D. Hauser, Inst. Pasteur, Unite des Toxines Microbiennes, 28, Rue du Docteur-Roux, F-75724 Paris Cedex 15,  
FRANCE

REFERENCE 2 (bases 1 to 4712)  
AUTHORS Hauser,D., Gbert,M., Eklund,M.W., Boquet,P. and Popoff,M.R..  
TITLE Comparative analysis of C3 and botlinal neurotoxin genes and their environment in Clostridium botulinum types C and D  
JOURNAL J. Bacteriol. 175 (22), 7260-7268 (1993)  
MEDLINE 94042900  
PubMed 8226673

FEATURES  
source Location/Qualifiers  
1..4712 /organism="Clostridium botulinum"  
/strain="type C (non toxic)"  
/sub\_strain="A028-CN page"  
/db\_xref="taxon:1491"  
791..796  
807..4682 /gene="neurotoxin mutated gene"  
807..4682 /gene="neurotoxin mutated gene"  
/pseudo  
/codon\_start=1  
/transl\_table=11  
/db\_xref="PID:e69305"  
2463..2465  
/misc\_feature /gene="neurotoxin mutated gene"  
/note="intrinsic stop codon"  
BASE COUNT 1907 a 482 c 721 g 1602 t



QY 1033 ATGAAGACGAGACCATGTACGCCGACACCACTCCACCGAGACATCTACGCCATCGGT 1092  
|||||  
Db 4353 ATGAAGATGAACATATGATGACATATCATAGTACTGAAGATATATATGCTATAGGT 4412  
|||||  
QY 1093 CTGGCTGACGACGACCAAGACATCAACGACACATCAATCTTCCAGATCCAGCCCAATGAAC 1152  
|||||  
Db 4413 TTAAGAGAACAAACAAGATATAAATGATATATATATTTCAATACACCAATGAAT 4472  
|||||  
QY 1153 AACACTTACTACTAGCTTCCCGAGATCTTCAAGTCCAACTCAACGGTGAACATCTCC 1212  
|||||  
Db 4473 AATACTTATATATGACATCTCAATATATTAATCAATATTTATGAGAGAAATATTTCT 4532  
|||||  
QY 1213 GGPATCTTCCATCGGTACCTACAGATTCGGTCTGCGTGGTGGTACTGGTACAGACACAAAC 1272  
|||||  
Db 4533 GGAATATGTTCAATAGTACTTATCGTTTATAGCTTGGAGGTGATGGTATAGACACAAAT 4592  
|||||  
QY 1273 TACTTGGTCCCACTGCTCAAGCAGGTACTACGCTCCTTCTGCTGGAGTCCACATCCACC 1332  
|||||  
Db 4593 TATTGGTGCTACTGTGAAGCAAGAAATATGCTTCAATATTAGAATCAACATCAACT 4652  
|||||  
QY 1333 CACTGGGATTCGTCGCCAGTCTCCCGAGTAA 1362  
|||||  
Db 4653 CATTGGGTTTGTACTCTGTAAGTGAATAA 4682  
|||||

RESULT 4  
CBCTOX  
LOCUS  
DEFINITION Botulinum bacteriophage genes for HA-17, HA-33, linear BCT 29-JUN-1999  
and C1 toxin.  
ACCESSION X62389  
VERSION X62389.1 GI:558175  
KEYWORDS botulinum toxin; C1 toxin; HA-17 gene; HA-33 gene; Hemagglutinin;  
neurotoxin; nontoxin; toxin.  
SOURCE Clostridium botulinum.  
ORGANISM Clostridium botulinum;  
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
Clostridium.  
REFERENCE 1 (bases 1 to 9613)  
AUTHORS Kimura,K., Fujii,N., Tsuzuki,K., Murakami,T., Indoh,T.,  
Yokosawa,N., Takeshi,K., Syuto,B. and Oguma,K.  
TITLE The complete nucleotide sequence of the gene coding for botulinum  
type C1 toxin in the C-ST phage genome  
JOURNAL Biochem. Biophys. Res. Commun. 171 (3), 1304-1311 (1990)  
MEDLINE 91024998  
PUBMED 2223445  
REFERENCE 2 (bases 1 to 9613)  
AUTHORS Tsuzuki,K., Kimura,K., Fujii,N., Yokosawa,N. and Oguma,K.  
TITLE Nucleotide sequence of the gene for one of the components of  
hemagglutinin produced by Clostridium botulinum type C  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 9613)  
AUTHORS Tsuzuki,K.  
TITLE Direct Submission  
JOURNAL Submitted (19-DEC-1991) K. Tsuzuki, Sapporo Medical College, Dept  
of Microbiology, South 1 West 17, Sapporo 060, JAPAN  
COMMENT On Oct 15, 1994 this sequence version replaced gi:40386.  
See also X53041 & D90210.  
FEATURES  
Location/Qualifiers  
1. 9613  
/organism="Clostridium botulinum"  
/strain="type C Stockholm, c-st"  
/db\_xref="taxon:1491"  
complement(111..551)  
/gene="HA-17"  
complement(111..551)  
/gene="HA-17"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA44260.1"  
/db\_xref="GI:40387"  
/db\_xref="SWISS-PROT:P46083"

/translation="MSSERTFLPNQYKIKSLFSDSLTYSSGALSFNTSLDNQK  
WKLVSISNGSFRSNAEPNKYLAVNDYGFIVLSSSSNNSLWNP\*KIANSYIICLT  
SIVNTDYAMTIYDNNNITDQPTLNPENFDINNSQILKLEK1"  
complement(560..565)  
/gene="HA-33"  
complement(614..1474)  
/gene="HA-33"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA44261.1"  
/db\_xref="GI:40388"  
/db\_xref="SWISS-PROT:P46084"  
/translation="MSOTNANDLRNNEVFIPSNNTNKVLDKISEVSEKVLNKLSGA  
NQKRLIYDNTNKQAYKIKVMDNTSLILITWAPLSSVSVKYDTGNDQYTWLLQNYLSR  
NRIIRYNMNPVLVQYIDDTLAVYQDTSSNQFFKFSNCIYEALNRNCKIQTLNS  
DRFLSKNLNSQIIVLQWQFSSRQKWIIEYNKFSAYTLKQENNYLFWIQNSYV  
ETVOSTSLIQWNIYLDNDASKYILYLDQTNRLVDVYNSQIANGTHVIVDSYHN  
TNQWMIINL1"  
complement(1482..1486)  
RBS 1726..1731  
RBS 1737..5327  
gene "nontoxin"  
1737..5327  
/gene="nontoxin"  
/codon\_start=1  
/transl\_table=11  
/product="non-toxic components"  
/protein\_id="CAA44262.1"  
/db\_xref="GI:40389"  
/db\_xref="SWISS-PROT:P46081"  
/translation="MDINDLINSPVDNKNVIVRAKNTIFFKAFKVPAPNIWVAPE  
RYGEPDIAEEYLDGIVDSNPLSODSERENFLOAIILLKRINNTSGKOLLSLI  
STAPPVGYIGGYSNPITFGKTQKSNKLSLVTSTIPRFGYRETNYIESON  
KNKFTASNIIFGGSNIVENNVIYKKNDAENGMAEIVFQPLTIYKNKYIDP  
AMELTKLILKLYGIDKPSDNLVVRRLTELDNKSQALIDLLISGGVDELEFI  
NLNPFYNTSYFNSIRKFEKKNYKTEEGNNAIGNDKRLKQKQINQVQDIWNL  
NLNPFQSFNSIIPRFSNALKHFYKQYTMQYTDNYNINNGVNGQINTKPLSKN  
TNIISKPKVVLVNNENNISLMKSNYDGLKGTEDFYTKYKPYNEEYEFNDS  
NPLNNTSIREVDSIPEIINDPKNSDNLVFTQITSMTEVTTHTALSINLQAI  
TNNEFTLSDFSKVSQKSLVYFSLDNLMSYLETKNDGPDIDTKYVILWKEVF  
KNYFDINTQELDSMGINBYLWFGKALNLTNSFVEEYQDSGALSLSKKNL  
REPIEDIDSLGUSFKDNLKLYEISKNIVYFKKIYFSLQWMTYYSQYFE  
LIMAKOSILAQESLVQIVONKFIKSIIPDILKILRETKTFFIDLSNESQIS  
MRYDNLNKAASICVFEVDIYKFIYMEYININNIKREFIORTNITKFNIDSKSILI  
NSYFKTIDFELDIQIKFNFQSQVQVMEKILSPYLLLFASKPNSNIIEIDSK  
NLQYTESIELYGVNGESLYLAKSNETIKFSNKFTNGLTINNTICFWRFTGK  
DKRLIGNKVNCGEIVFEDNGIVFRIIDSGNOESVYLSNIINDWYVITSVDRL  
KQDLIFINDKVNASIDOLISYTNILSLVKNKNSIYSELVLDNPLEITSEVIR  
NFFSLDYSYRDSKSLLEKNYQILNYVFFETSLYENDANKSYLSLKNYDGINI  
SSVFKLINDESKYVYOKWDECTICVDGTETKYLDISPENNRIQLVSSKDNKKITV  
NTDLFRPDCITFSYNDKYFSLSLRGGDYNMNCNDNNKVPKGAHLWILES"  
5330..5335  
5342..9217  
/gene="C1 toxin"  
5342..9217  
/gene="C1 toxin"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA44263.1"  
/db\_xref="GI:40390"  
/db\_xref="SWISS-PROT:P18640"  
/translation="MPITNNFYSDPVNKNILYDLTLNLANPEKAFRTGNW  
VIPDRFSNPNLKNKPRVSPKSGYDNPYLSLSDSKDTFLKEILKIFKRINSREI  
GEELIYRLSTDIPFGNNNTPTDFDVFNSVDVKTROGNNWKTGNSIPVILTG  
PRENIIDPETFTKLTNNFAAQGFALSIISIPRMLTYGNATNDVGEGRFSKE  
FCMDPILILMHELNEAMHLYGIAIPNDQITSSYNTSIFYSQYNVLEYAIVAFGGP  
TIDLIPKSAFYFEKALDYRSIAKRLNITTANPSSFNKYIGEKOKILIRYRVV  
ESSGEVYRNKVELYNELQJFEENYAKIYNQNRKLYLSNVTPTANTLDNIV  
YDIQGNFNPKNLNLVPMGONLSPALPKVPEPNMLYLFKECHKAIDGSLYKNT  
LDRELIVKNTDLDFIGDISDKTDFILRKDINEETETIYIPDNVSDQVILSKNTSE  
HGQLDLYPSIDSESLPGENQVYDNRQVYDNLNSYVLESQKLSNVEDFTTR  
SIEALDNSAKYVITYFTFLANKNAGVQGGFLMWANDVVEDEFTILNRKDLTKDIS

VSILPIVIGPALNISVRNGNETAFVGTVITILEEAPFETIPALGAFVIYSKVOE  
RNEIITIDNCLFORIKRKDSVEMWGMWLSRIITOPNNISYOMVDSIYQAGAIKA  
KIDLEFYKSSGSKENIKSOVENLNSLDVKISBAMNINIKFIRECSVTYLFKNMLPK  
VIDELNEEDNTAKLILIDSHNILLVEVDKLRKAKVNSPONTIPNIPFDFKLKSS  
LKIDINEFNINDSKILSLQRKNTLVDTGYNAEVBEEDQVLPNIPFDFKLKSS  
GEMRKVITQENIYNSWISFISFIRINKVSNLPVYIIDSXVNSGWSIGI  
ISNPLVETLQNEEDSQSINFSVDISNAPGYNKWFVTVNMMNKKIYINGKLID  
TIKVELTGINFSKTITFEINKIPDGLITSDNINMWIRDFIPAKELDKDINIL  
FNSLOVTVVVDYNGNDLYANKEYVMVNDILYRMYANSROIVENTRNNNDNEG  
KILIKRIGNINTRVRGGDILYFDMINKKAYNLFEMKNEYMADNHSHTEDIYALGLR  
EYKADINDIIFQIPMNNTIYASQIFSNFNGENISGICSGIRYRFLRGDMIRHN  
YLPTVKQNGYASLLESTHGWFPVSE"  
BASE COUNT 3846 a 1003 c 1348 g 3416 t  
ORIGIN

Query Match 48.58; Score 665.2; DB 1; Length 9613;  
Best Local Similarity 68.38; Pred. No. 1.4e-92;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
QY 13 ACCATCCCATTCACATCTTCTCCATACCAACCAATCCCTGTTGAAGGACATCATCAAC 72  
DB 7868 ACAATACCCCTTAATATTTTTCATATACATAAATCTTTATTAAGAGATATAATTAAT 7927  
QY 73 GAGTACTTCAACACATCAAGCATCCAGATCTGCTCCGTCGAGACCGTAAGACACC 132  
DB 7928 GAATATTCATATTAATTAATGATTCATTAATTTGAGCCCTCAACACAGAAAAATACT 7987  
QY 133 TTGGTCGACACCTCCGGTTACAACGCCGAGGCTCCGAGGAGGTTGACGCCAGCTGAAC 192  
DB 7988 TTAGTGGATACATCAGGATATAATGCAGAAAGTGTGAGTGAAGAGCGGATGTTCAAGCTTAAT 8047  
QY 193 CCAATCTTCCCATTCGACTTCAAGTGGTTCCTCCGTCGAGGACAGAGGTAAAGTCATC 252  
DB 8048 CCAATATTCCTATTTGACTTTAATAGTAGTGTTCAGGGGAGGATAGAGTTAAGTTATA 8107  
QY 253 GTACCCAGAACGAGAACATCGTTCACACATCCATGACGAGTCCCTTCCTCCTCCTC 312  
DB 8108 GTAACCCAGATGAATATGATATATAATCTATGATGAAGCTTTTAGCATTAGTTT 8167  
QY 313 TGATCAGATCAACACAGTGGTCTCCAACTGCGAGTTCACACATCATCGCTCCGTC 372  
DB 8168 TGATAGTAATAAATAAGTGAAGTAAATTTACCTGGATATACATAATATGATAGTT 8227  
QY 373 AAGAACAACCTCGGTGGTCCATCGGTATCATCTCCAACTTCCTGGTCTTCACCCCTGAAG 432  
DB 8228 AAAATAATCACTCAGGTGGAGTATAGTATATTAGTAAATTTTATGATTTACTTTAAAA 8287  
QY 433 CAGACGAGGACTCCGACGAGTCCATCAACTTCCTCAGACATCTCCACACAGCTCCT 492  
DB 8288 CAAATGAAGATAGTGAACAAAGTATAAATTTAGTTATGATATATCAATAATGCTCCT 8347  
QY 493 GGTACACAAGTGGTCTTCCTGTCACCTGTCACCAACACATGATGGTAAACATGAAGATC 552  
DB 8348 GGATACAATAATGCTTTTGTACTCTTACTAACAATATGATGGGAATATGAAGATT 8407  
QY 553 TACATCAACGGTAACTGATGACACCATCAAGTCAAGGAGTTGACCGGTATCACTTC 612  
DB 8408 TATATAATGGAATAATTAATAGATCTATAAAGTTAAAGTAACATGGAATTAATTTT 8467  
QY 613 TCAACACCATCACCTTCGAGATCAACAGATCCAGACACCGGTCTGTATCACTCCGAC 672  
DB 8468 AGCAAACTATAACATTTGAATAATAAATTCAGATACCGGTTTGTACTTCCAGAT 8527  
QY 673 TCGGACACATCAACATGTTGGATCCGTGACTTCATCATCTTCGCCAAGGAGTTGGACGGT 732  
DB 8528 TCTGATAACATCAATATGTTGGAATAGAGATTTTATATATTTGCTAAAGATTAAGATGGT 8587  
QY 733 AAGGACATCAACATCTGTTCAACTCTCTGAGTACACCAACGCTCGTCAAGGACTACTGG 792  
DB 8588 AAAGATATTAATATATTTTATAGCTTGCAATATCTATATGTTGTAAGATATTGG 8647  
QY 793 GGTACGACCTGAGATACAAACAGGAGTACATGTTGTTCAACATCTGACTTCTGAACAGA 852

DB 8648 GGAATGATTTAAGATATAATAAAGAATATTATATGTTAATATAGATTATTATAATAGA 8707  
QY 853 TACATGTACGCCACTCCACAGACATCGTCTTCAACACACAGAGGTAAACACACGACTTC 912  
DB 8708 TATATGTATGCGAACTCACCAGCAAAATGTTTTTAATACACGTAGAATAATTAATGACTTC 8767  
QY 913 AACGAGGTTTACAGATCATCATCAAGCGTATCAGAGGTAAACACACGACACAGAGTTC 972  
DB 8768 AATGAAGATATAAATATAATAAAGAAATCAGAGGAAATACAAATGATATAGTAGTA 8827  
QY 973 AGAGTGTGACATCTCTGACTTCGACATGACTATCAACACACAGGCTTACACCTGTTTC 1032  
DB 8828 CGAGGAGGATATTTTATATTTTGTATGACAAATTAATAACAAAGCATATATTTGTTT 8887  
QY 1033 ATGAAGAACGAGACCATGTACGGCGCAACACCTCCACCGAGGACATCTACGCCATCGGT 1092  
DB 8888 ATGAAGATGAACATATGTATGACAGATAATCATAGTACTGAAGATATATATGCTATAGGT 8947  
QY 1093 CTGGTGAGCAGACGAGGACATCAACGACACATCATCTTCCAGATCCAGCAATGAAC 1152  
DB 8948 TTAAGAAACAACAAGAGGATATAATTAATGATTAATATTTTCAATACACCAATGAAT 9007  
QY 1153 AACACTTACTACTACGCTTCCAGATCTTCAAGTCCCAACTTCAACGGTGAGACATCTCC 1212  
DB 9008 AATACTTATTTAGCATCTCAAAATTTTAATCAAAATTTTAAATGGAGAAAATATTTCT 9067  
QY 1213 GGTATCTGTTCCATCGGTACTACAGATTCCTGCTGGTGGTGACTGGTACAGACACAC 1272  
DB 9068 GGAATATGTTCAATAGTACTTATCGTTTATGACTTGGAGGTGATTGGTATAGACACAAT 9127  
QY 1273 TACTTGGTTTCAACGTCAAGCGGTAACTAGCCCTCTCTGCTGGAGTCCACTTCCACC 1332  
DB 9128 TATTTGGTGCCTACTGTGAAGCAAGAAATATGCTTCATTATTAGTAATCAACATCAACT 9187  
QY 1333 CACTGGGATTCGTCACGATCTCCGAGTAA 1362  
DB 9188 CATTTGGGTTTTGTACCTGTAAGTGAATA 9217

## RESULT 5

CBP1CNTC1 Clostridium botulinum phage 1C, CHN-14, CHN-33, CHN-138 and BONT/C1  
LOCUS 9689 bp DNA linear BCT 15-FEB-1999

## DEFINITION

genes.

ACCESSION X66433.1 GI:509271

VERSION X66433.1

KEYWORDS haemagglutinin; neurotoxin.

SOURCE Clostridium botulinum.

ORGANISM Clostridium botulinum

Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;

Clostridium.

REFERENCE 1 (bases 1 to 9689)

AUTHORS Hauser, D.F., Eklund, M.W. and Popoff, M.R.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 9689)

AUTHORS Hauser, D.F.

Direct Submission

Submitted (25-MAY-1992) D.F. Hauser, Institut Pasteur, Unite des

Toxines Microbiennes, Institut Pasteur, 28 rue du Docteur Roux,

F-75724 Paris Cedex, FRANCE

Location/Qualifiers

1. 9689

/organism="Clostridium botulinum"

/db\_xref="taxon:1491"

complement(253..647)

/gene="CHN-14"

complement(253..633)

/gene="CHN-14"

/codon\_start=1

/transl\_table=11

/product="14kDa protein associated with BONT

/C1-haemagglutinin complex"

/protein\_id="CAA47057.1"

## FEATURES

source

## gene

## CDS

```

/db_xref="GI:509272"
/db_translation="SPREMBL:Q45915"
/db_translation="MSSETEFLNGNKKLSPDSLYLYTSSGALSFSNTSSLDNOK
WKLEYISSSNGFESNAENKVLAYNDYGFVLSSSSSNLSNPKIALNSYIICTL
SIVNVYATYIINDNNHDLHF"
complement(642..647)
/gene="Chn-14"
complement(696..1556)
/gene="Chn-33"
complement(696..1556)
/gene="Chn-33"
/codon_start=1
/trans_table=1
/product="33kDa protein associated with BoNT
/C1-haemagglutinin complex"
/protein_id="CAA47058.1"
/db_xref="GI:509273"
/db_translation="MSTQANDLRNNEVFIPSSNNTKVLDKISOSEKLNKLSGA
NOKRLIYDYNKOAYKIKVMDNTSLITNNAPLSVSVKTDINGNOYKYLONYSR
NVIIRTNMNNLVLOYNIDTTLAVSTQTSNNOFFPSCNIEALNRNCKQOTQLNS
DRFLSKNLNQLIIVLWQWFDSSRQKWIETNFKSAYTLKQENNRILYLIQNSNNYV
ETYSTSLIQYWNLYLDNDASKYILYLNQDTRVLDVYNSQIANGTHIVYDVSYGH
TNQQWIIINLI"
complement(1564..1568)
/gene="Chn-33"
complement(1564..1568)
/gene="Chn-33"
join(1803..1808,1819..5409)
/gene="Chn-138"
1803..1808
/gene="Chn-138"
1819..5409
/gene="Chn-138"
/codon_start=1
/trans_table=1
/product="138kDa protein associated with BoNT
/C1-haemagglutinin complex"
/protein_id="CAA47059.1"
/db_xref="GI:4376225"
/db_translation="MDINDLNINSPVDNKNVIVIRKNTNFFKAFVAPNIWAPE
RYVEPDIAEYKLDGIVDSNPLSDSRENFLOAIILIKRINTTISGQLLSLI
STALFFYIGIGGSNPLFTFKTKPKSKKJNSLSTIPFPFGYRETNIEBSON
KNRYASNTIFGGGSNIVNNVYIKKNDANGMTMAEIVFPQLTYKRYNFIIDP
AMELTKLISLYLGIKPSDLNWPYRLTELDNKPQSLNIDILLISGGVLEFI
NTNYPWTNSFPNSIKMFYKNIYKTEEGNAIGNDIKRLKOKFOINQVDIWNL
NLVFCOSFNSIIPDRSNALKHRYKOYTYMDYTDYNGINGVNGCINIKPLSNKN
TNIISKPKVNLVNNENISLMKSNYIGDGLKGTEDFYSTYKIPYNEEYEFNDS
NPLNNTSIEFVDSIPELIDINPKKNSDLNVTQITSMEETHALSINYLQAI
TNNTFLSSDFSXVWSKOKSLVYSFLDNLNLSLETIKRNDGPDIDDKKIYLWLKEVF
KNYSPDINTQIEDSMGNEVILWFGKALNILNTNSFVEYQDSGALSLSKDDNL
REPIEDIDSLGLSFKDLNNKLYEISKNIVFKRYIFFLDQWTEYSQYFE
LICMAKQSLAQSILVQIKVQNKFTDLSKASIPDPLKIRETEKTFIDLNESQIS
MNRVDNFKASICVFVEDIYPKFISYMEKYINNINIKREFTORCNTINDNEKSILI
NSYTFKTDIREFLDIOIKNFNSOVQVMEKILSPVOLLFPASKGPSNILEDISGK
NTLLOYTESIELVGVNGESLYLKSNETIKESNKEFTGLTNNTFCWLFPEKND
DKTRLIGNKNVNGWEIYFEDNGLVFEIISNGNSVLSNINDNNWYISVDRL
KQDLILFNKNKNANSIDQLLSYTNIIISLVNKNNSIYVEELSVDNPIYSEEVIR
NYFSYLDNSYIRDSKSLLENKYNQLYNFFPETSLYEVDNKNKSLYLSKNTDGINI
SSVFKLINIDESYVYQKWDECIICVLDGTEKYLIDISPNRNRIQLVSSKDNAKITI
NTDLFRPDICITFSYNDKYFSLRDRDYNMINDCNKNKPKGAHLILES"
5408..5413
/gene="BoNT/C1"
5408..5413
/gene="BoNT/C1"
5424..9299
/gene="BoNT/C1"
5424..9299
/gene="BoNT/C1"
/codon_start=1
/trans_table=1
/product="Botulinical C1 neurotoxin"
/protein_id="CAA47060.1"
/db_xref="GI:509275"

```

```

/db_xref="SWISS-PROT:P18640"
/db_translation="MPTIINNFYSDPVDNKNLIYLDTHLNIANEPEKAFRITGNIW
VIPDRSRNPNLNKPPRVTSKPYDNPNTSLSDSDKDTLKEIILKLFKINSREI
GEELIVRLSDIIPGNNNTPIPTDFDVSVDVKTROGNNVKTGINSFVSITG
PRENIDPSTFKLINNTFAQEGGALSILISPREMLTYSNATNDWGEGRFSKSE
FCMDPILILMHELNHAMHNDIYGAIFNDQOTISSVTSNIYSQYNKLEVAEYIARGGP
TIDIPKSNRYFEERKALDYIRIAKRLNITTANPSSFNKYIGEKYKLIRKIRPVV
ESSEVTNRNKFEVLYNELTQITFEFNKAIYNQNRKYLNSNTPVPTANILDNV
YIQONGENIPKSNLNLFMGONSRNPAKRVNENMLYFTFKCHKADIGRSLNKT
LDCELLVKNTDLPFGDISDKTDFLARKDINEETEVIYVNDVSDQVILSKNTSE
HGOLDLPSIDSESEILPGENOVFYDNRQNDVYLSNYLLESQKLSQNDVEDFTFR
SIEBALNSAKVYTYPTLANKYAGVQSGFLMWANDVVEDETTNLRKDTLDKISD
VSAIPIGIPALNINSVRGNETEAFATGVITLLEAPEPTIPALGAFVYKVOE
RNEIITKIDMSQRKRWKSDYEMWMTGWSLRITQFNNSIQMTQDSLNLYAGAIKA
KIDLEYKYSKDKENIKSQVENLSDVKISEAMNNLNKPIRESVCTYFKNNMLPK
VIDELNEFDRTAKLINLDSHNIILVGEVDKLAKEVNSQNTIPFNIFYTNNSL
LKDIENFYNNINDSKILSLQNRKNTLVDTSYNAVESSEGDVQLNPIFFDFKLSS
GEDGRKVIYNNENIVNYSYESFISFNKYNKWSNLPGYTIIDSVKNNSSQWSIGI
ISNPLVTLAKQNSDESGSINFSYDISNAPGYNKFEVTVTNMGMNKIYINGKLID
TIKVKELGTNFSKTIITFEINKIPDGLITSDSDNINMKTROEYIFAKELDKDNLIL
FNSQYTNVYDIDWGNDRKYNKYVWVNDIYLNRYMYANSRQIVFNRNRNDFREGY
KILIRKIRGNTDTRVGGDILYFDMTINNKATNLFMKNETMADHSTEDYIATGLR
EQTKDINDNTIFQIQPMNNTYYASQIFKSNFNGENISGICISIGTYRFLGGDWYRHN
YLVPVTKQGNVYASLLESTSHWGFVPVSE"
9385..9424
/gene="BoNT/C1"
9385..9424
/gene="BoNT/C1"
/notes="putative"
BASE COUNT 3865 a 1022 c 1359 g 3443 t
ORIGIN
Query Match 48.5%; Score 665.2; DB 1; Length 9689;
Best Local Similarity 68.3%; Pred. No. 1.4e-92;
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;
QY 13 ACCATCCCATCAACATCTTCTCTACACCAACACTCCCTGTGTAAGGACATCATCAAC 72
Db 7950 ACAATACCCCTTAATATTTTTCATATATACTAATAATCTTTATTAAAGATATAATTAAT 8009
QY 73 GAGTACTTCACACATCAACGACTCCAGATCTCTGCTCGCAGAACCGTAAAGAACCC 132
Db 8010 GAATATTTCAATAATTAATTAATTAATTTTGAACCTTTTGACCCACAAAAGAAAATACT 8069
QY 133 TTGGTCGACACCTCCGGTACAAACCCGAGGCTCCCGAGGAGGCTGACGCTCCAGCTGAAC 192
Db 8070 TTAGTGGATACATCAGGATATAATTCACAACTGAGTGAAGAAGCGGATGTTACGCTTAAT 8129
QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCGGTGAGGACACAGGTAAAGTCAATC 252
Db 8130 CCAATATTTCCATTGACTTTAAATTAGGTAGTTAGTTCAGGGGAGGATAGAGGTAAAGTTATA 8189
QY 253 GTCACCCAGAACGAGAACATCGTCTTACAACTCCATGTACGAGTCTCTCTCATCTCTCTTC 312
Db 8190 GTAACCCAGAAATGAATATATGTATATAATTTCTATGTATGAAGTTTTAGCAATTAGTTT 8249
QY 313 TGGATCAGAAATCAAGTGGGTCTTCAACTTCCAGGTTCACACCATCATCGAGCTCCGCTC 372
Db 8250 TGGATTAGAAATAATAATGGGTAAAGTAAATTTTACCTGGATATACATAATGATAGTGT 8309
QY 373 AAGAACACTCCGGTGTGTCATCGGTATCATCTCCAACTTCCTCGGTCTTCCACCTCAAG 432
Db 8310 AAAATACTCAGTTGGAGTATAGGTATATTAGTAATTTTTTAGTATTATTCTTTTAAA 8369
QY 433 CAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCCTACGACATCTCCACACAGCTCCCT 492
Db 8370 CAAAATGAAGATAGTGAACAAAGTATAAAATTTTATGATGATATATCAAAATAATGCTCCT 8429
QY 493 GGTTCACAAAGTGTCTTCTCGTCACCGCTCCACACACATCATGGGTGATCAATCAAGATC 552
Db 8430 GGATACATAAATAGTGTGTGTGTAACCTTTTGTAACTGTTACTACATAATATGCGGAATATGAAGATT 8489
QY 553 TACATCAACGGTAAGCTGATCGACACCATCAAGGTCAAGAGGTTGACCGGTATCAACTTC 612

```

```
Db 8490 TATATAAATGGAATAATATAGACTATAAAGTTAAAGAACTAACTGGAATTAATTTT 8549
QY 613 TCCAGACCATCACTCGAGATCAAGAAGATCCAGACACCGGTCTGATCACTCCGAC 672
Db 8550 ACAGAACTATAACATTTGAATTAATAAATCCAGATACCGGTGATCTTCAGAT 8609
QY 673 TCCGACACATCAACATGCGATCGGAGCTTCACATCTCCGACGAGGTGGACGTT 732
Db 8610 TCTGATAACATCAATATGCTGATAAGAGATTTTATATATTTGCTAAAGAAATPAGATGT 8669
QY 733 RAGGACATCAACCTCTCTCAACTCTTCAGTACACCAACGTCGTCAGGACTACTG 792
Db 8670 AAGATATATATATATATATATATATATATATATATATATATATATATATATATAT 8729
QY 793 GGTACAGCTGAGATACACACAGAGTACTACATGTCACACGACTACTCTTGAACAGA 852
Db 8730 GGAATATATATATATATATATATATATATATATATATATATATATATATATATAT 8789
QY 853 TACATGTAGCGCACTCCAGACAGATGCTTCAACACGAGCTACACACGACTTC 912
Db 8790 TATATGTAGCGCACTCCAGCAATTTGTTTTATACAGTAGAATAATATATATATATAT 8849
QY 913 AACGAGGTTTCAAGATCATCATCAAGCGTATCAGAGGTATACACACGAGCTCAACCTGTC 972
Db 8850 AATGAGGATATAAATATATAAAGAAATCAGAGGAATACAAATGATAGTAGTA 8909
QY 973 AGAGGTGTGATCCTCTCTACATGCTACATGCTACACACAGGCTCAACACCTGTC 1032
Db 8910 CGAGGAGGAGATATTTATATTTGATGATGCAATTAATACAAAGCATATATTTGTTT 9869
QY 1033 ATGAGAAGGAGACCATGTAGCGCCAGACACCATCCACGAGGACATCTACGCCATCGGT 1092
Db 8970 ATGAAGANTGAACTATGTATGACAGATATCATATAGTACTGAAGATATATATGCTATAGT 9029
QY 1093 CTGCGTGACACACCAAGGATCAACACACATCATCTCCAGATCCAGCAATGAAC 1152
Db 9030 TTAAGAGAACAAAGAGATAAATGATATATATATATATATATATATATATATATATAT 9089
QY 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCC 1212
Db 9090 AATACTTATATATACGATCTCAATATTTAAATCAAAATTTAATGAGAAATATTTCT 9149
QY 1213 GGTATCTGTTCATCGGTACCTAGATATTCGTCGGTGGTGACTGGTACAGACACAAC 1272
Db 9150 GGAATATGTTTCAATAGGTACTTATCGTTTATAGACTTGGAGTGTATGATAGACAAAT 9209
QY 1273 TACTTGGTCCCACTGTCAAGAGGCTAACTACGCTCTCTGCTGGAGTCCACCTCCACC 1332
Db 9210 TATTGGTCCCTACTGTGAGCAAGAAATATATGCTTATTATAGATCAACATCAACT 9269
QY 1333 CACTGGGATTCGTCCTCCAGTCTCCGAGTAA 1362
Db 9270 CATTTGGGGTTTTGTACCTGTAAGTGAATAA 9299

RESULT 6
AB061780 11747 bp DNA linear BCT 06-NOV-2001
LOCUS Clostridium botulinum orf-22, ha-70, ha-17, ha-33, ntaha, nt genes
DEFINITION for ORF-22, HA-70, HA-17, HA-33, NTNHA, neurotoxin, complete cds.
ACCESSION AB061780
VERSION AB061780.1 GI:16580753
KEYWORDS
SOURCE Clostridium botulinum (strain:C-Voichi) DNA.
ORGANISM Clostridium botulinum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE 1
AUTHORS Sagane,Y., Kouguichi,H., Watanabe,T., Sunagawa,H., Inoue,K.,
Fujinaga,Y., Oguma,K. and Ohshima,T.
TITLE Role of C-terminal region of HA-33 component of botulinum toxin in
hemagglutination
```

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

Biochem. Biophys. Res. Commun. 288 (3), 650-657 (2001)  
21534265  
2 (bases 1 to 11747)  
Sagane,Y., Watanabe,T., Kouguichi,H., Morita,M., Miyamoto,T.,  
Yamamoto,T., Muroga,A. and Ohshima,T.  
Direct Submission  
Submitted (16-MAY-2001) Toshihiro Watanabe, Tokyo University of  
Agriculture, Faculty of Bioindustry, Yasaka 196, Abashiri, Hokkaido  
099-2493, Japan (E-mail:t-watanabe@bioindustry.nodai.ac.jp,  
Tel:81-152-48-3843, Fax:81-152-48-3843)  
Location/Qualifiers  
1. 11747  
/organism="Clostridium botulinum"  
/strain="C-Voichi"  
/db\_xref="taxon:1491"  
6. 545  
/gene="orf-22"  
/gene="orf-22"  
/codon\_start=1  
/transl\_table=11  
/product="ORF-22"  
/protein\_id="BAB71744.1"  
/db\_xref="GI:16580754"  
/translation="MNDLFYALENKHNDNQNFNFIEMSLKKYIKTSKKYNYDYNN  
DILYHLWKELIENLNFSELDLRKYSTIKRYCINICKKNDKRIIYNSEVYIK  
KLDVNVSYLCNFEFLDLISILNYKEQIIYMKFEGRKDNEAIRLSRSQSIYK  
IRUKSKLPIVMQLVNI"  
/complement(620..2491)  
/gene="ha-70"  
/complement(620..2491)  
/gene="ha-70"  
/codon\_start=1  
/transl\_table=11  
/product="HA-70"  
/protein\_id="BAB71745.1"  
/db\_xref="GI:16580755"  
/translation="MSLSIKELIYTKDSINNVLADGNVYVNRGDSWILSFQNLGLG  
GNISNGCTAIVGLRIRRETATPIYPTASFNEEYTKNNVQVNFANFEASIPGFE  
FSKTAPSKSLMYLQTYIRYEIKVLQNTVERAVLYVPSLVYVYKSTEFNSEQID  
KNFTYQSDKILNEKFIYKIDDTTVKESKNNNNNINFTSOTILPNSLYVINK  
GDGVMNDKDLIGTLIESSTSGSIQPLRNTTLPENTSNPTIFSOEYTEARLND  
AFQLENTSTTLFKFVEEATKNKISMKVNTYKELINTQNGINDKAPILPSL  
GKCEVSPSPQAPVETPVDDQGTQTPNENIIVGVINPSNIEE-STPIPDQDY  
NIPISQNNACYVLFKVTNTGVYKTKNNLPILIVEAIGSENNSNNISNDNIK  
AIRTITGNKSDAKSYLIVSLFKDKNYIRIPIQSSSTSQLFKRELGNISLDAST  
VNIIDLNTSGTHYTRQSPDVNGYISYQITPDGFNNFIASSIFSPTRNNQIGTLY  
RLTESINGYNIITNNYSDDLNVPEISLNGATYIFRVKVTLENNYNIIFDAIRNS"  
/complement(2505..2945)  
/gene="ha-17"  
/complement(2505..2945)  
/gene="ha-17"  
/codon\_start=1  
/transl\_table=11  
/product="HA-17"  
/protein\_id="BAB71746.1"  
/db\_xref="GI:16580756"  
/translation="MSSRTFLPNGYKIKSLFSNLIYLYSSGALSFSNTSLDNQK  
WKLEIVSSNGRFSNVAEPKNKYLAANDYGVLYLSSSSNNLWNPDKIANSYIICLT  
SIVNVDYATIIDNNNIIDQILNLPNFDINNQNQILKLEKL"  
/complement(2999..3856)  
/gene="ha-33"  
/complement(2999..3856)  
/gene="ha-33"  
/codon\_start=1  
/transl\_table=11  
/product="HA-33"  
/protein\_id="BAB71747.1"  
/db\_xref="GI:16580757"

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```

TYRTGNNQYWFNYLNNDSAMYTISNFSNQSKFLDVNSGLADGTNVQWVDSNGTS
AOKWIIIRL"
4119..7709
/gene="ntnha"
CDS
4119..7709
/gene="ntnha"
/codon_start=1
/transl_table=11
/product="NTNHA"
/protein_id="BAB71748.1"
/db_xref="GI:16580758"
/translation="MDINDDNINSPVDNKNVIVRARKNTNTFFKAFKVPAPNIWPAE
RYGPLEDIAEYGGYSDNIYDLSQDSRENFLOAIILLLKRINNTISGRKLSLI
STAFPPYGYGGYSGNIYFPGTKSNKLNLSVTSTPFPFGYRENTYIESQ
NMNFYASIVIFGPGSNVENNYIYKKNDNAENGMTMAEIVFQPLITYKKNFYIDP
AMELTKLISLFLYGIKPSDNLVWVRLRTELNDKOFQOLNIDLLISGGYDLFI
NNPWFNSVFPNSIKMFEXYKNYKTEIEGNAIGNDKLRKQFQINVOIDWNL
NLNYFQSEFNLIDPNSNALKHEKRYQYTMQYTDNYNINGVFNQINTEKLPNSKN
TNIIKPKVNLVNNNISLKNKNIYDGLKGTIEDFYSTYKIPNEYEYFNQSD
NPLNNISIEVDSEIPIIDINPYKDNSDLNVFTQITSMTEEVTHTALSINLQAI
NNENFTLSDFSKVYSKDSLYSFLDNLMSYLETIKNDGPIIDTKKYLWLKEVF
KNYSPDINTQIIDSMSGINEVWFGKALNINLNTSINSFVEEYQDSGALISLKKDL
REPNIIDIDISLGLSGFKDNLNKLVEIYSKNIVYFKIYFGLQDWTEYYSOYFE
LTCMAKQSLAQESVKQIVONKFTDLSKASIPDDTLKIRETEKTEFIDLSNESOIS
MNRVDFNLKASICVFVEDIYPKISVWEKXIYNNINIKTREFTORCININDNESLII
NSYTKIDFALDIQSITNFFNSQVQWKEILSPYQLLLFASKGNSNIIIEDSGK
NTLIDYTESIELVGVNGESLYLSPNETIKFSNKFTNGLTNNTNFTICFPLRETDNR
DTRLIGKNKNGWEIYFEDNGLVFIIDISNGQESVYLSNINDNMYIYSISVDR
KDQLIFINDKNVANSIDQLIYSYINILSVLNKNSIYVEELSDNPITSEEVIR
NYFVLDNYSIRDSKSLKYNLYNLYVFPETSLYEVDNKNKSVLSLKNQDGINI
SSVKPLNIDESKVVQWDECIICVLDTGTEKYLIDISPNRNRIQVSSKNKKTIV
NTDLFRPQIYFSYNDKYFSLSLRSDYNNMINDNKNKVPKGAHLWLES"
7724..11599
/gene="nt"
CDS
7724..11599
/gene="nt"
/codon_start=1
/transl_table=11
/product="neurotoxin"
/protein_id="BAB71749.1"
/db_xref="GI:16580759"
/translation="MPITINPNFSDPVDNKNILYDLTHLTLANEPEKAFRITGNW
VLPDRFSNPNKPNRPTSPKSYDNPYLSDDKDTFLKEIKLFRKINSREI
GELIYRSTDIPPPGNNTPIITDFDVEDVSDVKTROGNWNVKTSINPSVIITG
PENIIDPETSFKLTNTFAAQGFALSIIISPRFMTYSNANDVGEYFQFSKE
FCMDPILILHMLHANMELAGIAPNDOTISSVTSNIFYSOXNVKLEYAIEYFGSP
TIDLIPKARKYEKALDYRSIAKRLNLSITANPSPFNKYIGEYKQKILRYRW
ESSGVTVNRKKFVELYNELIQITERNIAKIYNQNKIYLSNVPITPNTAILDDNV
YDIQGFNIPKSNLVLPFGNLSRPNALRKVNPNENMLYLFTEKPAIDGRSLNKT
LDCRELLVNTDLFPIDIGISDVKTDFLRKDINTEEVYIYPDNVSDOVLKNTSE
HQLDLPLPSIDSEIILPGENQVYDNRNTQNDYLSYIYLESQKLSNVEDFTFR
STEEALDNSAKYIYFPTLANKVAGVGGFLMWANDVVEDFTNLRKDTLDKISD
VSAIIPYIGPALNITSVRGRNTEAPAVGTGVTILLEAFPEFTIPALGAFVIYSQOE
RNEIITDNCLEQRIKWKDSYEWGMWLSRIITOPNKSIXOMYDLSYQAQAIKA
KIDLEKYSQSDRENISQVENIKNSLDYKISAMANNINKFIRECSVYLFKNMPLK
VIDELNEPDRNKAHLNIDSHNIIIVGEVDKUKARKVNSFQNTIPNFISYTNLSL
LKDIINFEFNINDNSIKLSQNRKNTLDVTSYNAEVSDEGDVQLNIPFDFKLSSG
GDRGKVIYQNTENYNSMYSEFSISFWIRINKVNSLPGYTIIDSVKNNSWISGI
ISNVELTQKNESEQINFSDIISNAPGNKWFVTVMNMGMNKIYINGKILID
TIKVELTGKFSKTIITEINKIPDTGLTSDSDINMWRDRIYPAKELDKDINIL
FNSLOYVNVQWYDNGNDLRYNKEYVMYNDIYRYMANSROIVFNTRRNNDPNEG
KLIIRIRGNTNDRVRGDDILYFDMTINKKAYNLPKMKNETMADNHSDEDIYALGR
EOTKONINLIQIOPMNTIYISQIFKFSNFGENISGICSIQTYFRFLGGDWYRHN
YLPVTKQGNVYASLLESTSTHWGVPVSE"
BASE COUNT 4574 a 1263 g 1622 g 4288 t
ORIGIN

```

```

Query Match 48.5%; Score 665.2; DB 1; Length 11747;
Best Local Similarity 68.3%; Pred. No. 1.4e-92;
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;
QY 13 ACCATCCCATCAACATCTCTCTACACCAACAACTCCCTGTTGAAGGACATCATCAAC 72
|||||

```

```

Db 10250 ACAATACCCCTTTAATATATTTTTCATATATCTTAATAATCTTTATTAAAGATATATAAT 10309
QY 73 GAGTACTTCAACAACATCAACGACTCCAGACTCCTGTCCTCGAGAGCGTAAGAACACC 132
|||||
Db 10310 GAATATTTCAATAATATTAATGATTTCAAAAATTTTGACCTTACAAAACAGAAAAATACT 10369
QY 133 TTGGTCGACACCTCCGGTTTACAAACCCGAGGTCTCCGAGGAGGTGAGCTCCAGCTGAAC 192
|||||
Db 10370 TTAGTGGATACATCAGGATATTAATGCAGAAGTGAAGTGAAGGCGGATGTTTCAGCTTAAT 10429
QY 193 CCATCTTCCCATTCGACTTCAGCTCCGCTCCCTCCGCTGAGGAGGAGGTAAGGTGATC 252
|||||
Db 10430 CCAATATTTCCATTTGACTTTAAATAGGTAGTTCCAGGGGAGGATAGGTAAGTATA 10489
QY 253 GTCACCCAGAACGAGAACATCGCTTCAACCTCCATGTACGAGTCCCTTCCTCATCTCCTTC 312
|||||
Db 10490 GTAACCCCAATGAAAAATTTGTATATAATCTATGTAAGAAAGTTTGTAGCATTAGTTT 10549
QY 313 TGGATCAGAAATCAACAAGTGGGTCTCCAACTTGCAGGTTACACCATCATCGATCCGTC 372
|||||
Db 10550 TGGATTAGAAATAAATAATGGTAAAGTAAATTTTACCTGATATATACTATAATGATAGTGT 10609
QY 373 AGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACCTTCCTGCTTCCACCTCAAG 432
|||||
Db 10610 AAAAACTCAGTTGGTAGTAGTATATTAGTAATTTTGTAGTATTTTACTTTTAAA 10669
QY 433 CAGAAGGAGACTCCGAGCAGTCCATCAACTTCTCCTACGACATCTCOACAACGCTCCT 492
|||||
Db 10670 CAAAATGAAGATAGTGAACAAAAGTAAATTTTAGTTATGATATATCAAAATGAATGCTCCT 10729
QY 493 GGTTCACAAAGTGGTCTTCCTGTCACCGTCCACCAACAATGATGGGGTAACATGAAGATC 552
|||||
Db 10730 GGATCAATTAATGGTTTTTGTAACTGTACTAACATAATGATGGGAATATGAAGATT 10789
QY 553 TACATCAACGGTAAAGTGTATCGACACCATCAAGTCAAGGAGTTGACGGTATCAACTTC 612
|||||
Db 10790 TATATAATGGAATAATTAATAGATATATAAAGTTTAAAGAACTAACTGGTAATTAATTT 10849
QY 613 TCAAGACCATCACCTTCGAGATCAACAGATCCAGACACCGGCTGATCACTCCGAC 672
|||||
Db 10850 AGCAAACTATAACTTGAATAATAAATAATCCAGATACCGGTTGATTACTTACAT 10909
QY 673 TCCGAAACATCAACATGTGGATCGGTGACTTCTACATCTTCGCAAGGAGTTGAGCGGT 732
Db 10910 TCTGTAACATCAATATGTGGATAAGAGATTTTATATATTGCTAAAAGAAATAGATGGT 10969
QY 733 AAGGACATCAACATCCTGTTCACCTTGCAGTACACCAACGCTGCTCAAGGAGTACTGG 792
Db 10970 AAAGATTTAATATATATTTAATAGCTTTGCAATATATACTATGTTGTAAGAAATTTGG 11029
QY 793 GGTAAAGACTCAGATACAAACAGGAGTACTACATGGTCAACATCGACTACTTGAACAGA 852
Db 11030 GGAATGATTTAAGATATAATAAAGAAATTTATATGTTAATATAGATTAATTAATAGA 11089
QY 853 TACATGTACGCCAATCCAGACAGATCGCTTTCACACACGAGCTTAACAAACAGACTTC 912
Db 11090 TATATGTATGCGAATCCAGCAAAATGTTTAAATACCGTAGAATAATAAATGACATTC 11149
QY 913 AACGAGGGTTTACAAGATCATCATCAAGCGTATCAGAGGTAAACCAACAGACACAGATC 972
Db 11150 AATGAAGGATATAAATAATAAAGAAATCAAGAGAAATACAAATGATAGATA 11209
QY 973 AGAGGTGGTACATCTCTGACTTCAGATGATCAACACAGGCTTACAACTGTC 1032
Db 11210 CGAGGAGGAGATTTTATATTTTGTATGACAATAATAAACAAGCATATAATTTGTTT 11269
QY 1033 ATGAAGAAGCAGACCATGTACCGCCGACACCACTCCACCGAGGAGATCTAGCCATCGGT 1092
Db 11270 ATGAAATGAACATATGTATCGAGATAATCATAGTACTGAGATATATATGCTATAGT 11329
QY 1093 CTCGGTGAACAGCAAGGAGATCAACGACAAACATCATCTTCCAGATCCAGCCATGAAC 1152
Db 11330 TTAAGAGAAACAAACAAAGGATATAAATGATATAATATTATTTTCAATACAAACCAATGAAT 11389

```

```

QY 1153 AACACTTACTAGCTCCAGATCTCAAGTCCAACTTCAACGGTGAGAACATCTCC 1212
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 11390 RATACTTATTATGAGATCTCAATAATTATTAATCAAAATTTTATGGAGAAATATTCT 11449
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1213 GGTATCTTCTCCATCGGTACAGATTCGGTCTGGTGGTGGTACGACGACACAC 1272
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 11450 GGAATATGTTCAATAGGTACTATTCGTTTATAGACTTGGAGGTGATTGGTATAGACACAAT 11509
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1273 TACTTGGTTCACATGTCAACGAGGTAAGTACGCCCTTGTCTGGAGTCCACTCCACC 1332
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 11510 TATTGGTGCTACGTGACGACGAGGAATATGCTTCATATTAGATCAACATCAACT 11569
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
QY 1333 CACTGGGATTGCCCCAGTCTCCGAGTAA 1362
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
Db 11570 CATTTGGGTTTGTACCTGTAAGTGAATAA 11599
  || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||

RESULT 7
LOCUS CECBONT
DEFINITION Clostridium botulinum C phage BONT/C1, ANTP-139, ANTP-33, ANTP-17,
  ANTP-70 genes and ORF-22.
ACCESSION X72793.1 GI:516171
VERSION Antp gene; bont gene; DNA binding protein; neurotoxin type C1.
KEYWORDS Clostridium botulinum phase 1C.
SOURCE Clostridium botulinum phase 1C.
ORGANISM Viruses.
REFERENCE 1 (bases 1 to 12297)
  Hauser,D., Eklund,M.W., Boquet,P. and Popoff,M.R.
  Organization of the botulinum neurotoxin C1 gene and its associated
  non-toxic protein genes in Clostridium botulinum C 468
  Mol. Gen. Genet. 243 (5), 631-640 (1994)
  94301293
  MEDLINE 8028579
  PUBMED
REFERENCE 2 (bases 1 to 12296)
  Hauser,D.F.
  Direct Submission
  Submitted (03-MAY-1993) D.F. Hauser, Inst. Pasteur, Unite des
  Toxines Microbiennes, 28, Rue du Docteur Roux F-75724, Paris Cedex
  15, FRANCE
  On Jul 27, 1994 this sequence version replaced gi:509268.
  Related sequences: X53751 & X66433.
FEATURES
  source
    1..12297
      /organism="Clostridium botulinum phase 1C"
      /strain="468"
      /sub_species="type C"
      /db_xref="taxon:35348"
      /lab_host="Clostridium botulinum"
      289..294
        /note="ORF22"
      302..841
        /note="ORF22"
      /codon_start=1
      /transl_table=11
      /protein_id="CAA51308.1"
      /db_xref="GI:509269"
      /db_xref="SPTREMBL:Q38195"
      /translation="MNDLFVAIENLKHDNCHFNFIEMSLKKVYIETSKYNLYDYNN
        DILVHLAKELIENLKNFSELDLRKYISTSKRYCINICKKRDKKIYNSEVYK
        KLDVANYYSICDNFELDLSLTKYKQIYMKFEGRKDNETAIRLSRQSIYK
        IRTISLKLPIVQNLVNI"
      complement(854)..2787
        /gene="ANTP-70"
      854..906
        /note="putative; ORF22"
      complement(854)..906
        /gene="ANTP-70"
      /note="putative"
      complement(916)..2787
        /gene="ANTP-70"

gene
terminator
terminator
CDS

```

```

/codon_start=1
/transl_table=11
/protein_id="CAA51309.1"
/db_xref="GI:509270"
/db_xref="SWISS-PROT:P46085"
/translation="MELSILKELYITKDKSINNVLADGNVYVNRGDNILSRQNLG
  GNSNGCTAIVGDDLRITETATPYPTASNEEYIKNNVQNVFANFTEASEIPIGE
  FSTAPENSKSLMYLQYTYIRYEIIKVLQNTVERAVLYPSLGVKSIETSESEID
  KNEPFTSQDKILNEFYIKKIDDTITVKESKNNNINFNFTSQTILPPLGLYVINK
  GDGYMTNDKDLIGTLIBSSSGGIOPRLNNTPLFNTSNPTIFSQETEARLND
  AFNIOLENTSTLFLKEVEAPTNKISMVNTYKVELINVQNGNIDDKAEYVLP
  GKCEVSDAPSPQAPVETPDQDGIQGPENIIVGVINSENFEEISTPIPDYTY
  NIPTIGNACIVLFVNITGVYKTIITKNLPLELIIEAIGSSNNMNSNLSDNIK
  AIKIITGLARSDAKSLIVSLFKDNKYIRIPISSTTSQIFKRELGNTSLDASIT
  VNILDNLTSGTHYTRQSPDVGNYISQLTIPGDFNNIASSIFSRFRNNOGIGTLY
  RITESINGYNLITINNYSOLLNVEPISLINGATYIFRVKYTELANNYIIFDAYRNS"
  complement(3801)..3252
    /gene="ANTP-17"
  complement(3801)..3241
    /gene="ANTP-17"
/codon_start=1
/transl_table=11
/protein_id="CAA51310.1"
/db_xref="GI:516172"
/db_xref="SWISS-PROT:P46083"
/translation="MSSRTPLPNGNYKIKLSFSDSLTYTSSGALSFSNTSSLDNOK
  WKLEYISSNGRFSNVAEPNKYLAINDYGFYLSSSNNSLWNPDKIAINSYIICTL
  SIVNVDIAWTIIDNNNIIDQPIINLPNFDINNNSQILKLEK"
  complement(3247)..3252
    /gene="ANTP-17"
  complement(3304)..4176
    /gene="ANTP-33"
  complement(3304)..4164
    /gene="ANTP-33"
/codon_start=1
/transl_table=11
/protein_id="CAA51311.1"
/db_xref="GI:516173"
/db_xref="SPTREMBL:Q38196"
/translation="MSQNTANDLRNNEVFFISPSNNTNKVLDKISQSEVKLWNKLSGA
  NOKRWLIYNTNQAYKIKYMDNTSLILTNAPLSSVSVKTDNGDNQVYLLQNVISR
  NVIIRNMPNLVLOINHDDTLVSTQISSNQFEKFNCIYEALNENCKLOLQNS
  DRFLSNLNASQIIVLMQWEDSSKQWIIYEINETRSAYLKCQENNRNLTWQNSNRY
  ETYQSDSLGIQYWNINDASKYILYLNQDTRNRLVDVYNSQIANGTHIVYDSFHGN
  TNOQWNLNI"
  complement(4171)..4176
    /gene="ANTP-33"
  4313..8017
    /gene="ANTP-139"
  4313..8017
    /gene="ANTP-139"
  4411..4416
    /gene="ANTP-139"
  4427..8017
    /gene="ANTP-139"
/codon_start=1
/transl_table=11
/protein_id="CAA51312.1"
/db_xref="GI:516174"
/db_xref="SPTREMBL:Q38197"
/translation="MDINDLANINSPVDNKNVYIVRAKNTNFFKAFKVPANWVAPE
  RYVEPLDIAEYKLDGGIYDSNFIQSODSERENFLQAIILLKRNTNISGQLLSLI
  STALPYPYVIGGGYSSPNLFTGKTPKSNKLSLVTSTIPFPGGYRETNITESON
  NKPFYASNVIIFPGSNTIVENNVIYKKNDANGMTAEIVFOPLLYKYNKVIDP
  AMELTKLIKSLFFLIQIRPSNLYVVPKRLTELONKQPSQLNIDLLISGVDFLEFI
  NTNPFYNTSYFPNSIKMFEKYKIKYTEIGNNAIGNDKLRLKQKQFQINQDILWNL
  NLNYFQGSNFIIPDRFSNALKFYKQYITMDYTDYTNINGFNQNGTINLPNSKN
  NLNISKPEKVNVLNENNISLAKSNIIYDGLGTTEDFYSTYKIPYNEIYRNDSD
  NPLNNISIEVDSIPEDINPDKNSDLVFTQITSEVTTHTALSINLQAOI
  TNNENFTLSSDSKSVYSSKSLVSLYSLDNLSLEYLTKNQDPIDTKDYVWLKEVF
  KNYSPDLNLTQEDSDMCGINEVAFEGKALNLTNSFVEYQDSGALSISKDNL
  REPNIEDDISLGLSFDNLNKLIEIYSKNIYFKKIFSFELDQWTEYYSQYFE
  LICMAKQSLAQESLYKQIVQNKFTDLSKASIPDPDTLKLIIRRETTTEKTFIDLSNESQIS

```

```

MNRVDFNLKASICVFVEDIYKPFISYMEKYINNIINIKTRFQIORTCTNINDNEKSLII
NSYFKTIDRFKDIOISIKNFENSOVEOMKEILSPYOLLFPASKGPNINIEDISGK
NTLQYTESLELVANGESLYLKSPNETIKFSNKFTNGLNNEFTICPWFLEFTKND
KTLIANKYNNKNGWEIYFEDNGLYFEIIDSNGNQESVYLSIINDWNWYIISVDR
KQDLIINDNANWISDIILSIYISLNNKNSIYVEELSVDNPIIISSEVIR
NYFSLDNIYROSSKSLLEKNYOLYNYVPEPTELYEVNDRNKSYSLSKNTDGINI
SSVFKLINDSEKSYQKWDCEICVLDGTEYLDISPENNRIOIVSKSDNAKRTIV
NIDLFRPCDITFSYNDKYFSLSLRDGDNMCMICNDNNKVPKGAWHLWILES"
7931..12032
/gene="BONT/C1"
/mrna ">12032
7931..8021
/gene="BONT/C1"
8032..11907
/gene="BONT/C1"
/codon_start=1
/transl_table=11
/protein_id="CAA51313.1"
/db_xref="GI:516175"
/translation="MPTIINFNYSDPVNKNLYLDTHLNLANPEKAFRITGNIW
VIPDRSRNSPNLNKPRVTSKPSGYDPNLYSTDSDKDTFLKEIKLFPKRINSREI
GEELIYRLSTDIPPGNNTPINTDFEDVFNVDVKTROGNNWVKTGSGINSVITG
PRENIIDPESTFKLTNTTAQEGFALSISISPREMLTYSNATNDYEGREFSKE
FCMDPILLLHMLNHMHNLYGAIENDQIISVTSNIFYSQYNWLEVAEYARGGP
TIDLPKASRYFEKALDIYETSTAKRLNSITIANPSSFNKYIGETQKQLIKRYRPV
ESSEGVNRRKFEALDYNEITQIFTEFYAKIYNQNRKIYLSNVPTVYANILDDNV
YDIONGPNKSNLNFMGONLSRNPALRKVPENMLYFTKFCHEKAIIDGRSLYNT
LDCRELVAKNTDFEIGDISDKTDFLRKIDINEETEVYIPDNVSDQVILSKNTSE
HGOLDLVPDSSESLIPGENOVFYDNRTQNYDLYNSVYLESQKLSNDBDFPTR
SIBERLONSADVTFEPLANKYNAGVCGGLEFLMWANDVVEDETTNLRKDLKDISD
VSAIPIYGPALNINSVRGNTEAFVAVGVVILLAEPEFIPALGAFYIYSKVQE
ARELITKDNCLEORIKRKQDSVEMMGTWLSRIITQFNISITQMSDLYNLQAGALKA
KIDLEYKISGSDENKSVENLKNLSLDKISSEAMNNINKFIRECSVILNKLMLPK
VIDELNEPDRNTAKLNLIDSHNLIILVGEVDKLKAKVNSFQNTIPFNFISYNNLS
LKDINEFNNDINSKILSLQNRKNTLVDTSYNAEVSESDQVLPNIPFPDFKLGSS
GDRGVKVTQONENIVNSWVESFISFWIRINKYNKWFVTVNNMGNMKIYINGKLID
ISNLVFTPLKONENIVNSWVESFISFWIRINKYNKWFVTVNNMGNMKIYINGKLID
TIKVKELGTFNFKTIFELNKIPDTGLIPSDSDNINMLRDFYIFAKELDGDKNIL
FNSQYVNYVDYNGNOLRYNKEYYVNYIDILRYNYANSRQIVFVIRNRNNDPNRQY
KIIIRKRGNTNDRVRGGDILYFDMTINNKAYNLFMKNMETIYDNHSEDIYAGLR
EQTQKIDNDNIIFQPKMNTYIYASQIFKSNFNGENISGICSGTYFRFLGGDWYRHN
YLVPVTKQGNVTSALLESSTSHWGFVPVSE"
11993..12032
/terminator
/gene="BONT/C1"
BASE COUNT 4825 a 1287 c 1703 g 4482 t
ORIGIN
Query Match 48.5%; Score 665.2; DB 7; Length 12297;
Best Local Similarity 68.3%; Pred. No. 1.4e-92;
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;
QY 13 ACCATCCCATCAACATCTTCTCCACACCACTCCCTGTTGAAGGACATCATCAAC 72
Db 10558 ACAATACCTTTAATATTTTTCATATCTAATAATATCTTTAATAAAGATATAAT 10617
QY 73 GAGTACTTCAACAACATCAACGACTCCAGAPCCGTCCCTCGAGAACCCGTAAGAAC 132
Db 10618 GAATATTTCAATAATATTATGATTCAAAAATTTTGAGCCTACAAAACAGAAAAA 10677
QY 133 TTGGTGAGACCTCCGGTTACACGCCGAGGTCCTCCGAGGAGGTGAGCTCCAGCTGAC 192
Db 10678 TTAGTGATACATCAGGATATATTCAGAAGTGAGTGAGAGGCGGATGTTTCAGCTTA 10737
QY 193 CCAATCTCCCATTCGACTTCAAGCTGGGTTCCCTCCGTTGAGGACAGAGGTAAGTCA 252
Db 10738 CCAATATTTCCATTTGACTTTAAATTAGTAGTCTCAGGGGAGGATAGAGGTAAAG 10797
QY 253 GTACCCAGACAGACATCGTCTACACTCCATGTCAGGAGCTCTTCTCCATCTCCTTC 312
Db 10798 GTAACCCAGAGTAAATATTGTATATAATCTATGATGAAGTTTATAGCATTAGTTT 10857
```

```

QY 313 TGGATCAGAATCAACAAGTGGGTCTCCAACTTTCGAGGTTACACCATCATCGACTCG 372
Db 10858 TGGATTAGAATAAATAAATGGTAACTTACCTGATATATAATAATGATAGTGT 10917
QY 373 AGAACAACACTCGGTGGTTCCTCGGTATCATCTCCAACTTCTCGGTCTTCACCCG 432
Db 10918 AAAATAACTCAGGTGGAGTATAGGTATATTAGTAAATTTTATAGTATTACTTTAA 10977
QY 433 CAGAAGCAGGAGTCTCGGACAGTCCATCAACTTCTCTACGACATCTCCAAACAAC 492
Db 10978 CAAATGAAGTAGTGACAAAGTATAAATTTTAGTTATGATATCAATAATGCTCT 11037
QY 493 GGTTCACAACAAGTGGTCTTCGTCAACCGTCAACCAACATGATGGGTGATCAAGATC 552
Db 11038 GGATACAATAAATGGTCTTTTGTAACTGTTTACTAACAATATGATGGGAAATATGA 11097
QY 553 TACATCAACGGTGAAGTGTATCGACACCATCAAGGTCAAGGAGTTCACCGGTATCA 612
Db 11098 TATATAATGGAATAATTAATAGATCTATAAAGTTAAAGACTAACTGGAATAATTT 11157
QY 613 TCCAAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCTGTGATCACTCCG 672
Db 11158 AGCAAAACTATAAACAATTTGAAATAAATAAATTCAGATACCGGTTTGATTACTCAG 11217
QY 673 TCCGACACATCAACATCTGATCGGTGACTTCTACATCTTCGCCAAGGAGTTGGAG 732
Db 11218 TCTGATACATCAATATGATGGTAAGAGATTTTATATTTTGTCTAAGAAATTAGATGG 11277
QY 733 AAGGACATCAACATCCTGTTCAACTCTTCGAGTACCAACGTCGTCGCAAGGACTACT 792
Db 11278 AAAGATATTAATATATTATTAATAGCTTGCATATACTAATGTTGTAAAGATATTGG 11337
QY 793 GGTACGACCTGAGATCAACAAGGAGTACTACATGCTCAACATCGACTACTTGAACAG 852
Db 11338 GGAATGATTTAAGATAATAAAGAGATATTATATGTTAATAGATATTATTAATAGA 11397
QY 853 TACATGTACGCAACTCCAGACAGATCGTCTTCAACACACGACGATCAACAACGACTTC 912
Db 11398 TATATGTACGCAACTCCAGACAAATTTCTTTTAATACACGTCAGAAATAATAATG 11457
QY 913 AACGAGGTTACAAAGATCATCATCAAGGTATCAGAGGTACACCAACGACACGAGTC 972
Db 11458 AATGAAGGATATAAATAATATAAATAAAGATCAGAGGAAATACAAATGATAGTAGA 11517
QY 973 AGAGGTGTGACATCCTGCTACTTCGACATGACTATCAACAACAAGGCTTCAACCTG 1032
Db 11518 CGAGGAGGAGATATTTTATATTTTATGATGACAAATTAATAACAAGCATATAATT 11577
QY 1033 ATGAAGACGAGACATGTACGCCGACAAACCACTCCACCGAGGACATCTACGCCATG 1092
Db 11578 ATGAAGATGAACATGTATGTCAGATATATCATAGTACTGAAGATATATATGCTATAG 11637
QY 1093 CTGCGGTGAGCAGCAAGGACATCAACGACAAACATCATCTTCCAGATCCAGCCAA 1152
Db 11638 TTAAGAGACAAACAAGGATATAATGATATAATATATTTCAATACAACTAATGAAT 11697
QY 1153 ACACHTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAGAACATCTCC 1212
Db 11698 AATACTTATTATTACGCACTCAAAATATTAAATCAAAATTTTAATGAGAGAAATATTCT 11757
QY 1213 GGTATCTGTTCCTCGGTACCTACAGATTCCGCTCGGTGCTGCTGCTGCTGCTGCTG 1272
Db 11758 GGAATATGTTCAATAGGTACTTATCGTTTTAGACTTTGAGGTGATGGTATAGACAA 11817
QY 1273 TACTTGGTTCCAACTGTCAAGCAGGTAACTACGCTCTCTTCTGCTGGAGTCCACTTCA 1332
Db 11818 TATTTGGTCTCTACTGTGAAGCAAGGAAATTTATGCTTCATTATTAGATCAACATCACT 11877
QY 1333 CACTGGGGGATTCGTCCTCCAGTCTCCGAGTAA 1362
Db 11878 CATGGGGTTTTGTACCTGTGAAGTGAATAA 11907
```



```

RESULT 8
AB037920      11631 bp      DNA      linear      BCT 24-JAN-2002
LOCUS         Clostridium botulinum genes for ORF-22, HA-70, HA-17, HA-33, NTNHA,
DEFINITION   neurotoxin, complete cds.
ACCESSION    AB037920
VERSION      GI:6939789
KEYWORDS     neurotoxin; NTNHA; HA-33; HA-17; HA-70; ORF-22.
SOURCE       Clostridium botulinum (strain:D-4947) DNA.
ORGANISM     Clostridium botulinum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
REFERENCE    1
AUTHORS      Kouguchi,H., Watanabe,T., Sagane,Y., Sunagawa,H. and Ohyama,T.
TITLE        In vitro reconstruction of the Clostridium botulinum type D
              progenitor toxin
JOURNAL      J. Biol. Chem. 277 (4), 2650-2656 (2002);
MEDLINE      21659747
REFERENCE    2 (bases 1 to 11631)
AUTHORS      Sagane,Y., Watanabe,T., Kouguchi,H., Yamamoto,T., Takizawa,J.,
              Kawabe,T., Murakami,F., Muroga,A., Nakatsuka,M. and Ohyama,T.
TITLE        Direct Submission
JOURNAL      Submitted (01-FEB-2000) Tohru Ohyama, Tokyo University of
              Agriculture, Faculty of Bioindustry; Yasaka 196, Abashiri, Hokkaido
              099-2493, Japan (E-mail:t-oyama@bioindustry.nodai.ac.jp,
              Tel:81-152-48-3838(ex.368), Fax:81-152-48-2940)
FEATURES     Location/Qualifiers
             1..11631
                /organism="Clostridium botulinum"
                /strain="D-4947"
                /db_xref="taxon:1491"
             6..545
                /gene="orf-22"
             6..545
                /gene="orf-22"
                /codon_start=1
                /transl_table=11
                /product="ORF-22"
                /protein_id="BAA90656.1"
                /db_xref="GI:6939790"
                /translation="MDLFDVAENLKHNDHDFDIEMLSKKYIEKTSKYNLYDYNN
                DILHMKELFEINLKNPSELDRKYISIRKRYCINCKKRNDRKKIYINSEATYK
                KLEAVNYSCEDEFELDTLSILYKKEQIYKPFCKRDNKEIARRLHLRSQSYK
                IRIKSLKLYPIVMQLVNI"
             complement(621..2492)
                /gene="ha-70"
             complement(621..2492)
                /gene="ha-70"
                /codon_start=1
                /transl_table=11
                /product="HA-70"
                /db_xref="GI:6939791"
                /translation="MSLSIKELYTKDKSINNVLADGNVYVNRGDSWILSRQNMIG
                GNISNGCTAVGLRIRREYATPYVYPTAFNDEYIRNNVQVFNFTASRPIGFE
                FSKTAPSKNGIYNLQTYIRYKELIKVRNIVIRAVLYVPSLGAKSFNSGGDID
                ANGYTSDXCKILNEKFIYKKAETTTAKESDNNNTNLNTSQTILPYNGLYVINK
                GDNVTRNDKDLGILLIENYSGSIQPRNTRPLRNTSPTFLSQEYTEARLND
                AFNIOLENTSTLFEVVEAPDNKNIEMKAYNTYERKLYNQNGNADKAEYIPLS
                KCESYDAPSPQAPVPTVQDGFIOIGPNENIIVGVNPSENIEISITPIPDYTY
                NIPISQNNACVLYFTVNTGVYKINAQNLPLLIYESIGSDMMIQNTLSNNIK
                ALNVTGDSNARSALIVSLIKKNXYIRIPOLSSSTNOLIFKRELGNISDLANST
                VNILDNLNTSHTHYTQSDPVGNVYSQITIPGDNNNIASSIFSTRNNGOIGHLY
                RLTESIGNYNTIKNTSDLLNVPISLLNGATYIFRVKVTNLNINIIIFAIRNS"
             complement(2506..2946)
                /gene="ha-17"
             complement(2506..2946)
                /gene="ha-17"
                /codon_start=1
                /transl_table=11
                /product="HA-17"
                /protein_id="BAA90658.1"
                /db_xref="GI:6939792"
             complement(3008..3868)
                /gene="ha-33"
             complement(3008..3868)
                /gene="ha-33"
                /codon_start=1
                /transl_table=11
                /product="HA-33"
                /protein_id="BAA90659.1"
                /db_xref="GI:6939793"
                /translation="MSQTNANDLRNNEVEFFISPSNNTNKVLDKISGSEVKLNKLSGA
                NQWRLLYDNKQAYTIKVMNDTSLITDAPLSSVSVAKTIDNINNQVLIQDRIISR
                NVILRNMNPLVQNTDITLIVTQTNNOQFFKFSNCIYEALNRRCKIQTOLNS
                DRFLSKNLNSQIIVLQWDFDSSRQKWTIEYNETKSAYTLKCOENRRLYLNNSYV
                ETYQSTDLSIQYWNINYLNDASKYILYNLQDTRVLDVYNSQTANGTHVIVDSYHGN
                TNOQMIINLI"
             4129..7719
                /gene="ntaha"
             4129..7719
                /gene="ntaha"
                /codon_start=1
                /transl_table=11
                /product="NTNHA"
                /protein_id="BAA90660.1"
                /db_xref="GI:6939794"
                /translation="MDINDDLNINSVDNKNVIVFARKNTFFFAKVPAPNIWAPE
                RYXGEDLIAEYKLDGGIYDSFLSODSERENFLQAIITLKRINNTISGKQLLSLI
                STAIPFYGVGGYSPNITFTGKPKGNKKILNSVTIPIPFPGGRETNYIESQN
                KNFVASNIVIFGSGNIVENNVCYKKNDAENGMTMAEILFQPLLAYKKNFYIDP
                AMELTKLISLFLYGLGVVYRLRTLDELNKOFSQNLIDLLISGVDLEFI
                NTNPYFTNSYFSNISKMEKYKNIYETIEGNAIGNDKLRLKQKQNSQVDLWNL
                NLNFSKEPNSIIPDRFSNALKHYKQYTDYGDNYANGVANGQNTKLPLSDKN
                TNIISKPEVNSLVNANNISKSNYIGDKGTEDFYSTIKLIPNEYEYFRNDS
                NFLENNTSIEVDSIPEIDINPYKNSDDLLFTQITTEVITHALPVNVLQAI
                ITNENFTLSDFSKVSSKDLVYSLDNLMSYLETIKNDGPIPTDKYILWLKEVF
                KNYSEIDNLTQEDISGQINEVWFGKALNLTNSFVEYQSGPISLISKDNL
                SEPNEIDDIPDLSLGSFVKLNLKLYETYSKRVRYFYFNFLDQWTEYYSOYFE
                LICMAKOSILQASVCVVEDIYPKFIYSMEKYINNINIKREIORTNINDNEKSIIS
                NRYVDFNLKASICVFVEDIYPKFIYSMEKYINNINIKREIORTNINDNEKSIIS
                NSYFTKIDPKINQIAIKNFNFQSVQVQVMEKLSYQALLFATGCPNSNIEDISGK
                NTLQYTESVELYGVNGESLYLKSPEVTEVFNFSNFTNGLTNFTICFWLFTSGDD
                DKTLIGKNCNCGWEIYFDNGLVFEIISDNGQESVYLSVNINNNWYIISVDRLL
                KQDLIFINDKNVANYSIEQILNLYTNVLSVKNNSYVEELSVDLPVASEVIR
                NYFSYLDNSYIRDSKSLLEYKNLYNYFPETSLYEVDNNSKSLKMTDGINI
                PSVKFLINDESKGVQKWDECIIGVSDGTEKYLDISPENNRIOLVSSKDNNAKITV
                NTDLPKPCITFTSYNDKYFSLSLRDGDYNMIMCNDNNKVPKGAHLWLAS"
             7734..11591
                /gene="nt"
             7734..11591
                /gene="nt"
                /codon_start=1
                /transl_table=11
                /product="neurotoxin"
                /protein_id="BAA90661.1"
                /db_xref="GI:6939795"
                /translation="MTWPVKDFNYSDFVNDNDILYRIPONKLIITTPVKAFTONINW
                VIPERFSDSNLSLSPKPRPTSKYQYDPSTYSLTDEQDFTLGLIKLIFKRNERDI
                GKRLINLYVSGPMSGDSPTDFDTRHTTNIAVEKFENGSKVYNTIITPSVLIFG
                PLPNILDTASLQGOOSNPSPEFGFTLSILKVAPEFLTFDSVTNQSASVGLKSI
                FCMDDVIALMHELTSHOLYGINIPSDKRIPOVSEGFSDQGNVQPEELTYFGGS
                DVELIQLERLQREKALGHYKIDAKRLANNIKTIPSSWSNIDKYKIFSEKYNFDK
                YTINGFNVLITKGFINSQNIERNPALKQSSVDLFTKVLCLLRNRSRDSSTC
                IQVKNNTLPYADKDSISEIFESQITIDETNVNYSDFNPSLDESIDARVITNPEAV
                DPLLPNTYMEPLNVPGEVEFYDDITKDVDLYNSYILEAQLSNNVENITLTVEE
                ALGYSNKIYTFPLSLAEKVNKGQAGFLNWAENEVEDFTNIMKDTLDKISDVSAI
                IPYIGPALNTGNSALRGNFKAATAGVAFLEGFPEFTIPALGVFTFYSSQREKRI
                IKTENCLQREKVKWSDYOMVSNMLSRITTFRNHISYQWDSLSYQADAKAKIDL
                EYKYSQSDRENTKSQVENLKSNDKXISEAMNNINKFIRECSVTYLFKNMLPKVIDE
                LNKFDLTKTLELINLDSHNIILVGEVDRLKAKVNSFENITFFNIFSTINNSLAKDM
                INEYFNSINDSKILSLQNKNTLMDTSGYNAEVRVEGNVOLNPFIFFDFKLGSSGDDR

```



GVIVTQNEIVNAMYSEFISFWIRINKWYNLPYTIIDSVKNSGWSIGIISNF  
LFTLKQNSQEDINFSYISKNAAGYKWFVITTTNMGNMMIYIINGKIDIKV  
KELTGINFSTIIFOMNKIDRITSDSDNIWMDYIFAKELDDKIDNILENSL  
QYTNVVDYGNLDRYDKEYIMLVNMYNMYKSGNGIVENTRNKNDNFNGYKII  
KRIGTNDTRVGENVLYPNTIDNKQYSLGKXPSNLGTDLPGLNLDPMDEIR  
KYGSIIFOPNTEDYASQFLPSNATIRLGIISGSYFKLGDDIWFNHEYLIPVI  
KIEHYASLESSTHWFVPASE"  
BASE COUNT 4493 a 1254 c 1657 g 4227 t  
ORIGIN

Query Match 32.8%; Score 449.8; DB 1; Length 11631;  
Best Local Similarity 59.38; Pred. No. 1.3e-59;  
Matches 807; Conservative 0; Mismatches 542; Indels 12; Gaps 2;  
QY 13 ACCATCCCATTAACATCTTCTCTACACCAACATCCCTGTTGAAGGACATCATCAAC 72  
DB 10248 ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTAATAAAGATATGATTAAT 10307  
QY 73 GAGTACTTCAACATCAACAGCATCCAGATCCTGCCGTGAGACCGCTAAGACACC 132  
DB 10308 GAATATTCAATAGTATTATGATTCAAAAATTTTGAGCTTACAAAATAAAAAATACT 10367  
QY 133 TTGGTGCACCTCCGGTTACAACGCGAGCTCCGAGGAGGTGACGTCACGCTGAAC 192  
DB 10368 TTGATGGATACATCAGGATATAACGACGAGTGAAGTGAAGGCAATGTTCACTTAAT 10427  
QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGTTCCTCCGGTGAGGACAGGTAGGTCATC 252  
DB 10428 CCAATTTTCCATTTGACTTTAAATAGGTAGTTCAGGGGATGATAGAGTAAAGTTATA 10487  
QY 253 GTCAACGAGAGAGACATCGTCTACAACATCCATGATGACGAGTCCCTTCTCCATCTCCTC 312  
DB 10488 GTAACCCAGAAATGAAATATTGTATATATGCTATGATGAAGTTTTAGTATTAGTTT 10547  
QY 313 TGGATCAGATCAACAGTGGTCTCCAACTTCCAGGTTACACCATCATCGACTCCGTC 372  
DB 10548 TGGATTAGGATAAATAATGGTGAAGTAAATTTACCTCGATATACTATAATGATAGTT 10607  
QY 373 AAGAACAACCTCCGGTTGGTCCATCGGTATCATCTCCAACTCCCGTTCACCCCTGAAG 432  
DB 10608 AAAATAACTCAGGTGGAGTATAGTATTATTAGTAAATTTTGTAGTTTACTTTAAAA 10667  
QY 433 CAGACGAGGACTCCGAGGATCCCATCAACTTCCTCAGCATCTCCACACAGCTCCT 492  
DB 10668 CAAATGAAATPAGTGAACAAGATATAAATCTTASTTATGATATATCAAAAGATGCTGG 10727  
QY 493 GGTTACAACAAGTGGTCTTCCGTCACCGTCCACCAACATGATGGGTAAACATGAAGATC 552  
DB 10728 GGATATAAATAGTGTTTTGTACTATTACTACCAATATGATGGAATATGATGATT 10787  
QY 553 TCATCAACGGTGAAGCTGATGACACCATCAAGGTCAAGGTGACCGGTATCACTTC 612  
DB 10788 TATATAAATGGAATTAATAGATATACTATAAAGTTAAAGAGTTAACTGGAATTAATTT 10847  
QY 613 TCCAAGACCATCACCTTCGAGATCAACAAGATCCAGACACCGGTCTGATCACCTCCGAC 672  
DB 10848 AGCAAACTTATACATTTTCAATGAATAAATTCCAATATCTGCTTAATTACCTCAGAT 10907  
QY 673 TCGACAAACATCAACATGGGATCGGTGACATTCCTCAGCATCTCCACAGGAGTGGAGGT 732  
DB 10908 TCTGATAACATCAATATGGGATAGGATTTTATATCTTTTGTCTAAAGATTAATGATGAT 10967  
QY 733 AAGGACATCAACATCTGTTCACTCCCTGAGTACACCAACGCTCGTCAAGGACTACTGG 792  
DB 10968 AAAGATATTAATATATTTATTTATAGCTTGCAATATACTATAATGTTGTAAGATATTGG 11027  
QY 793 GGTAAAGCCTGAGATACAAACAAGGAGTACATGTTGGTCAACATCGACTACTTGAACAGA 852  
DB 11028 GGAATGATTTAAGATATGATAAAGATATTACATGATTAACTGTAATTAATGATGATA 11087  
QY 853 TACATCTAGCCCACTCCAGACAGATCTCTTCAACACCAACGCTTACACACGACTTC 912  
DB 11088 TATATGCTCTAAAAAAGGCAATGGAAATGCTTTTATACACGCTAAAAATAATAGACTTC 11147

QY 913 AACGAGGTTTAAAGATCATCATCAAGCGGTATCAGAGTTAACCAACGACACGAGATC 972  
DB 11148 AATGAAGGATATAAATTTATAATAAAGAAATAGAGGAAATCAAAATGATAGATA 11207  
QY 973 AGAGTGTGACATCTCTGCTACTTCGACATGACTATCAACACACAGGCTTACACCTGTC 1032  
DB 11208 CGAGGAGAAATGATTTATATTTTAACTACATAATGATACAAACAATATAGTTAGGT 11267  
QY 1033 ATGAAGACGAGACATGATACGCGCAACACCATCTCCACCGAGGACATCTACGCCATCGGT 1092  
DB 11268 ATGATAAACCTTCTA-----GAAATCTAGGAGTCTATTAGTTCCACTAGGTGCA 11318  
QY 1093 CTGCGTGAAGACAGCAAGGACATCAACGACACATCATCTTCAGATCCAGCAACGAAC 1152  
DB 11319 TTGGATCAACCAATGGATGAGATACGTTAAATATGTTGCTTTTATAATACACCATGCAAT 11378  
QY 1153 AACACTTACTACTCGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCC 1212  
DB 11379 ACTTTTGTATTACTGATCATCAATTTTGTCAAGTATGCAACACAATAGGCTT 11438  
QY 1213 GGTATCTTTCATCGGTACCTACAGATTCGGTCTGGTGGTGAC---TGTACACACAC 1269  
DB 11439 GGAATPACTATCAATTTGGTATGTTAGTTTCAAACTTGGAGATGACTATTGGTTAATCAC 11498  
QY 1270 AACTACTTGGTTCCAACCTGTCACGAGGTAACTACGCTTCTGCTGGAGTCCACTTCC 1329  
DB 11499 GAAATTTAATCTCTGTATATAAATAGAGCATTTATCTTCATTTATAGATCAACATCA 11558  
QY 1330 ACCCATCGGATTCGTCCTCAGTCTCCGAGTAAATAGGAAT 1370  
DB 11559 ACTCATTTGGTGTGTACCTGCAAGTGAATAATAAATAGT 11599  
RESULT 9  
CLOTNCO 4049 bp DNA linear BCT 18-DEC-2001  
LOCUS Clostridium botulinum type D gene for neurotoxin consisting of  
DEFINITION botulinum neurotoxin D and Cl, complete cds.  
ACCESSION D38442  
VERSION D38442.1 GI:1374775  
KEYWORDS neurotoxin consisting of botulinum neurotoxin D and Cl; neurotoxin.  
SOURCE Clostridium botulinum D (strain:South African) DNA.  
ORGANISM Clostridium botulinum D  
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.  
REFERENCE 1 Moriishi, K., Koura, M., Abe, N., Fujii, N., Fujinaga, Y., Inoue, K. and Ogumad, K.  
Mosaic structures of neurotoxins produced from Clostridium botulinum types C and D organisms Biochim. Biophys. Acta 1307 (2), 123-126 (1996)  
AUTHORS 2 (bases 1 to 4049)  
Moriishi, K.  
TITLE Direct Submission  
JOURNAL Submitted (29-SEP-1994) Kohji Moriishi, Osaka University, Research Institute for Microbial Diseases, Research Center for Emerging Infectious Diseases; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kohji@biken.osaka-u.ac.jp, Tel:81-6-6879-8343, Fax:81-6-6879-8269)  
FEATURES  
source 1. 4049  
/organism="Clostridium botulinum D"  
/strain="South African"  
/db\_xref="taxon:56829"  
163..4020  
/codon\_start=1  
/transl\_table=11  
/product="neurotoxin consisting of botulinum neurotoxin D and Cl"  
/protein\_id="BA007477.1"  
/db\_xref="GI:1374776"

/translation="MTWVPKDFNYSDPVNDNDILYLRIPONKLIITPVKAFMITQNIW  
VIPERFSDNPSPRPRTSKYISYDFSYLSDEQDFTLKGILKFLKRENERDI  
GKLIYLVVSPFMDSDSPSTDFTRHTNIAVEKPEFGSKVNTIITPSVLFIG  
PLNILDYASLQGOOSNPSPEGTLSILKVAPEFLITDFSTNOSAVLKGSI  
FCMDPVIALMHELTSHSLHOLYGINPSDKRIPOVSEGFSDSGPNVPOPERLYTFGSK  
DVEIIOERLQREXALGHVYKDIARLKNINKTIPSSMSNIDKVKKIFPSKYNFKD  
DNTGNFLVNDKNSLSLDITVMSVIVSYSQYNKRNHYFSKHYLPVFNILDDNI  
ITYINGNLITIKGFNENSGNLERPALQKLSSSVVDLFTKVLRLTRNRDSDTC  
IOVANNLPVADKDSIOBESQITIDETNVNSDNFSDLESILDAKVPNPEAV  
DPLPNYMPFLANVPGEVEFYDDITKVDYLSNYYLEAQLNSNVENITITTSVEE  
ALGYSNRIYTFPSLAEKVKNGVQAGFLNNAVEVDEFTNIMKRDLDKLSVSAI  
IPYIPALNTGNSLRNFKOATAGVAFLEGFPEFTIPALGVTFVPSIOERKII  
IKTIGLEORGNRWRKDSYOMYSNWLIRITQFNHISYOMYDSLSYQADAIAKIDL  
EYKYSDDKENTKSVENLKNLSIDVKISEAMNNINKFIREGSVYLFKNMPLPKVIDE  
LNKEDATKLELINDHNLIDVEVRLKARKVNESFENTIPFNIFSTINSLSKDM  
INFEYNSINDSKILSONKNTLMDTSGYNAEVNGLNPFPFDFKLGSDDDR  
GLIIVTONENINAMVESISFWIRINKXNVSFLTPTIIDSXVKNNSGWSIGIISNF  
LVFTLKONENISQDINFYSIDSKNAGYKWFVTITITNMGNMIYINKGLIDIKY  
KELTGINFSTIITFQNNKIPNTGLITSDSDINMWIRDFYIFAKELDDKIDINILNSL  
KYTVVVDYNGDLRLDYKEYIMLVNVMNRYMSKKGIVFTRKNNDFN9GYKII  
KIRGNTIDRVGEENVLYFNTIDNKQYSLGWYKPSNLTGDLPLGALDQDMDLIR  
KYGSFILOPNTEDYASQFLSSNATNRLGLISIGSYSPKLGDDYFNHLYLPIVI  
KIEHYALSLESTSHWVVFPASE"

BASE COUNT 1621 a 438 c 630 g 1360 t

Query Match 32.7%; Score 448.2; DB 1; Length 4049;  
Best Local Similarity 59.2%; Pred. No. 2.6e-59;  
Matches 806; Conservative 0; Mismatches 543; Indels 12; Gaps 2;

QY 13 ACATCCCATTCATCTCTTACACCAACATCCCTGTTGAAGGACATCATCAAC 72  
DB 2677 ACAATACCCCTTTAATATTTTCATATACTAATAATCTTTATTAAGAAGATATGATTAAT 2736  
QY 73 GAGTACTTCAACAACATCAAGACTCCCAAGATCCTGCCCTGCAGACCGCTAAGAACC 132  
DB 2737 GAATATTCATAGTATTAAATGATTCAAAATTTTGACITACAAAATAAATAAATACT 2796  
QY 133 TTGGTCGACCTCCGGTTACAAACCCGAGGTCTCCGAGGAGGTGACGTCACGCTGAAC 192  
DB 2797 TTGATGGATACATCAGATATAAACGAGAACTGAGATAGAAAGCAATGTTCAGCTTAAT 2856  
QY 193 CCAATCTCCCATTCGACTTCAAGCTGAGTCTCCGTCGAGGACAGAGTAAAGTCATC 252  
DB 2857 CCAATATTCATTTGACTTTAAATAGTAGTAGTTCAGGGATGATAGAGTAAATATTA 2916  
QY 253 GTACCCAGAACGAGAACATCGTCTACAACTCCATGACGAGTCTCTCCATCTCCTTC 312  
DB 2917 GTAACCCAGAAATGTAATATGCTATGATGAAGCTTTTAGTATTAGTTTT 2976  
QY 313 TGGATCAGAAATCAAGTGGTCTCCAACTTCCAGGTTACACCATCATCGACTCCGTC 372  
DB 2977 TGGATTAGGATAAATAAGTGGTAAAGTAAATTTACCTGGATATATAATGATAGTGT 3036  
QY 373 AGAACAACCTCCGGTTGTCATCGTATCATCTCCAACTTCCGTCCTTCCACCTGAAG 432  
DB 3037 AAAAAATCACTAGGTGGAGTATAGTATTATAGTAATTTTGTAGTTTACTTTAAAA 3096  
QY 433 CAGAAGCAGGATCCGAGGATCCATCACTTCTCCATGACATCTCCACACCGCTCCT 492  
DB 3097 CAAATGAAATAGTGAACAGATATAAATCTTATGATATATCAAGAATGCTGGG 3156  
QY 493 GGTACACAAGTGGTCTCTCCCTCCCTCCACCAACATGATGGGTAAACATGAAGATC 552  
DB 3157 GGAATATAAATAGTGTCTTCTAATTTACTATTAATACCAATATGATGGGAATATGATGATT 3216  
QY 553 TACATCAACGGTAAGTGTATCGACACCATCAAGGTCAAGGATGTACCGGTATCAACTTC 612  
DB 3217 TATATAATGGAAATTAATAGATATAAAGTTAAAGATTTAACTGGAATTAATTTT 3276  
QY 613 TCAACAGCATCATCCTCGAGATCAACAAGATCCAGACACCGGTCTGATCACCTCCGAC 672  
DB 3277 AGCAAAACTATACATTTCAATGATGAATAAATTCCAATATCTGCTTAATTACCTCGAT 3336

RESULT 10  
AC117900/c  
LOCUS  
DEFINITION Rattus norvegicus clone CH230-320C16, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 33 unordered pieces.  
ACCESSION AC117900  
VERSION AC117900.3 GI:21746813  
KEYWORDS HTG: HTGS PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 151705)  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Avele, M., Banks, T.,  
Barbaria, J., Benton, J., Bimege, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Bowie, S., Burlett, C., Burrell, K.L., Byrd, N.C.,  
Carroll, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,





```
----- Project Information
Center project name: GfrI
Center clone name: CH240-358N16
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 120888 bases at least Q40
Consensus quality: 127504 bases at least Q30
Consensus quality: 132322 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
      (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 61 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.
  *
  1 1019: contig of 1019 bp in length
  * 1020 1119: gap of unknown length
  * 1120 2779: contig of 1660 bp in length
  * 2780 2879: gap of unknown length
  * 2880 4496: contig of 1617 bp in length
  * 4497 4596: gap of unknown length
  * 4597 5621: contig of 1025 bp in length
  * 5622 5721: gap of unknown length
  * 5722 6889: contig of 1168 bp in length
  * 6890 6989: gap of unknown length
  * 6990 8124: contig of 1135 bp in length
  * 8125 8225: gap of unknown length
  * 8226 9904: contig of 1580 bp in length
  * 9905 11392: contig of 1488 bp in length
  * 11393 11493: gap of unknown length
  * 11494 12676: contig of 1184 bp in length
  * 12677 12777: gap of unknown length
  * 12778 13861: contig of 1085 bp in length
  * 13862 13962: gap of unknown length
  * 13963 15708: contig of 1746 bp in length
  * 15709 17582: contig of 1775 bp in length
  * 17583 17682: gap of unknown length
  * 17683 19005: contig of 1323 bp in length
  * 19006 19105: gap of unknown length
  * 19106 21071: contig of 1966 bp in length
  * 21072 21171: gap of unknown length
  * 21172 22855: contig of 1684 bp in length
  * 22856 22955: gap of unknown length
  * 22956 24423: contig of 1468 bp in length
  * 24424 24523: gap of unknown length
  * 24524 26272: contig of 1749 bp in length
  * 26273 26372: gap of unknown length
  * 26373 28549: contig of 2177 bp in length
  * 28550 28649: gap of unknown length
  * 28650 30139: contig of 1490 bp in length
  * 30140 30239: gap of unknown length
  * 30240 31817: contig of 1578 bp in length
  * 31818 33917: gap of unknown length
  * 33918 33762: contig of 1845 bp in length
  * 33763 33862: gap of unknown length
  * 33863 36122: contig of 2260 bp in length
  * 36123 36222: gap of unknown length
  * 36223 38477: contig of 2255 bp in length
  * 38478 38577: gap of unknown length
  * 38578 40590: contig of 2013 bp in length
  * 40591 40690: gap of unknown length
  * 40691 43154: contig of 2464 bp in length
  * 43155 43254: gap of unknown length
  * 43255 45512: contig of 2258 bp in length
  * 45513 45612: gap of unknown length
```

```

  * 45613 47325: contig of 1713 bp in length
  * 47326 47425: gap of unknown length
  * 47426 49849: contig of 2424 bp in length
  * 49850 49949: gap of unknown length
  * 49950 51626: contig of 1677 bp in length
  * 51627 51726: gap of unknown length
  * 51727 52957: contig of 1131 bp in length
  * 52958 52957: gap of unknown length
  * 52959 56168: contig of 3211 bp in length
  * 56169 56268: gap of unknown length
  * 56269 59010: contig of 2742 bp in length
  * 59011 59110: gap of unknown length
  * 59111 61204: contig of 2094 bp in length
  * 61205 61304: gap of unknown length
  * 61305 63328: contig of 2024 bp in length
  * 63329 63428: gap of unknown length
  * 63429 66088: contig of 2660 bp in length
  * 66089 66188: gap of unknown length
  * 66189 67828: contig of 1640 bp in length
  * 67829 67928: gap of unknown length
  * 67929 70649: contig of 2721 bp in length
  * 70650 70749: gap of unknown length
  * 70750 72883: contig of 2134 bp in length
  * 72884 72983: gap of unknown length
  * 72984 75071: contig of 2088 bp in length
  * 75072 75171: gap of unknown length
  * 75172 78234: contig of 3063 bp in length
  * 78235 78334: gap of unknown length
  * 78335 80048: contig of 1714 bp in length
  * 80049 80148: gap of unknown length
  * 80149 83463: contig of 3115 bp in length
  * 83464 83633: gap of unknown length
  * 83634 85719: contig of 2356 bp in length
  * 85720 85819: gap of unknown length
  * 85820 87592: contig of 1773 bp in length
  * 87593 87692: gap of unknown length
  * 87693 91127: contig of 3435 bp in length
  * 91128 91227: gap of unknown length
  * 91228 94240: contig of 3013 bp in length
  * 94241 94340: gap of unknown length
  * 94341 98820: contig of 4480 bp in length
  * 98821 98920: gap of unknown length
  * 98921 104013: contig of 5093 bp in length
  * 104014 104113: gap of unknown length
  * 104114 107600: contig of 3487 bp in length
  * 107601 107700: gap of unknown length
  * 107701 111658: contig of 3958 bp in length
  * 111659 111758: gap of unknown length
  * 111759 115121: contig of 3363 bp in length
  * 115122 115221: gap of unknown length
  * 115222 120413: contig of 5192 bp in length
  * 120414 120513: gap of unknown length
  * 120514 125376: contig of 4862 bp in length
  * 125376 125475: gap of unknown length
  * 125476 129676: contig of 4201 bp in length
  *
  Query Match      13.7%; Score 188.4; DB 2; Length 176351;
  Best Local Similarity 49.6%; Pred. No. 9.3e-20;
  Matches 483; Conservative 0; Mismatches 491; Indels 0; Gaps 0;

  QY 236 ACAGAGTAAGGTCATCGTCACCCAGAGAGAGATCGTCTACAACTCCATGTGACGAGT 295
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  Db 52708 ACAGACACAGCAGCAGCAACACACAGCAACACAAACAAACACACACAGCA 52649

  QY 296 CCCTTCCTCCTCTCTGATCAGATCAACAAGTGGGTCTCCAACTGGCCAGGTTCAC 355
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  Db 52648 ACAACACACACACACAGCAGCAGCAACACACACAGCAGCAGCAACACACACA 52589

  QY 356 CCATCATCGACTCCGGTCAAGAGAACTCCGGTTGGTCCATCGGTATCATCTCCAACTCC 415
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  Db 52588 GCAGCAACACACACAGCAGCAACACACAGTCTTTTATACAGCAGCAACACACA 52529

  QY 416 TGGTCTTCACCCCTGAAGCAGACAGGAGCTCCGAGCAGTCCATCATCTCTCTAGCACA 475
```



\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

```

1 1195: contig of 1195 bp in length
* 1196 1295: gap of unknown length
* 1296 2301: contig of 1006 bp in length
* 2302 2401: gap of unknown length
* 2402 3597: contig of 1196 bp in length
* 3598 3697: gap of unknown length
* 3698 4705: contig of 1008 bp in length
* 4706 4805: gap of unknown length
* 4806 6053: contig of 1248 bp in length
* 6054 6153: gap of unknown length
* 6154 7709: contig of 1556 bp in length
* 7710 7809: gap of unknown length
* 7810 9075: contig of 1266 bp in length
* 9076 9175: gap of unknown length
* 9176 10387: contig of 1212 bp in length
* 10388 10487: gap of unknown length
* 10488 12171: contig of 1684 bp in length
* 12172 12271: gap of unknown length
* 12272 13923: contig of 1852 bp in length
* 13924 14233: gap of unknown length
* 14234 15253: contig of 1230 bp in length
* 15254 15353: gap of unknown length
* 15354 17056: contig of 1703 bp in length
* 17057 17156: gap of unknown length
* 17157 18630: contig of 1474 bp in length
* 18631 18730: gap of unknown length
* 18731 20564: contig of 1834 bp in length
* 20565 20664: gap of unknown length
* 20665 22920: contig of 2256 bp in length
* 22921 23020: gap of unknown length
* 23021 24042: contig of 1022 bp in length
* 24043 24142: gap of unknown length
* 24143 25407: contig of 1265 bp in length
* 25408 25507: gap of unknown length
* 25508 27326: contig of 1819 bp in length
* 27327 27426: gap of unknown length
* 27427 28925: contig of 1399 bp in length
* 28926 28925: gap of unknown length
* 28926 30295: contig of 1370 bp in length
* 30296 30395: gap of unknown length
* 30396 32901: contig of 2506 bp in length
* 32902 33001: gap of unknown length
* 33002 35383: contig of 2382 bp in length
* 35384 35483: gap of unknown length
* 35484 38276: contig of 2793 bp in length
* 38277 38376: gap of unknown length
* 38377 39868: contig of 1492 bp in length
* 39869 39968: gap of unknown length
* 39969 41970: contig of 2001 bp in length
* 41970 42069: gap of unknown length
* 42070 44382: contig of 2313 bp in length
* 44383 44482: gap of unknown length
* 44483 47592: contig of 3110 bp in length
* 47593 47692: gap of unknown length
* 47693 52151: contig of 4459 bp in length
* 52152 52251: gap of unknown length
* 52252 56760: contig of 4509 bp in length
* 56761 56860: gap of unknown length
* 56861 61177: contig of 4317 bp in length
* 61178 61277: gap of unknown length
* 61278 66148: contig of 4871 bp in length
* 66149 66248: gap of unknown length
* 66249 71886: contig of 5338 bp in length
* 71887 71987: gap of unknown length
* 71988 78079: contig of 6093 bp in length
* 78080 78179: gap of unknown length
* 78180 81154: contig of 2975 bp in length
* 81155 81254: gap of unknown length

```

```

* 81255 88890: contig of 7636 bp in length
* 88891 88990: gap of unknown length
* 88991 97276: contig of 8286 bp in length
* 97277 97376: gap of unknown length
* 97377 107376: contig of 10000 bp in length
* 107377 107476: gap of unknown length
* 107477 116754: contig of 9278 bp in length
* 116755 116854: gap of unknown length
* 116855 127329: contig of 10475 bp in length
* 127330 127429: gap of unknown length
* 127430 136746: contig of 9317 bp in length
* 136747 136846: gap of unknown length
* 136847 153947: contig of 17101 bp in length
* 153948 154047: gap of unknown length
* 154048 169163: contig of 15116 bp in length.
FEATURES
  source
    1. 169163
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-261G15"
BASE COUNT 41981 a 37857 c 38661 g 43240 t 7424 others
ORIGIN
Query Match 13.3%; Score 183; DB 2; Length 169163;
Best Local Similarity 50.6%; Pred. No. 6.3e-19;
Matches 469; Conservative 0; Mismatches 455; Indels 3; Gaps 1;
QY 233 AGACAGAGGTAAAGTTCATCGTCACCCAGACGAGCAACATCGTTACAACTCCCATGTACG 292
DB 35132 AAGCCAAACCAACACACACAGCAGCAACAAACAAACAGCAACACAGCAGCA 35073
QY 293 AGTCCTTCTCCATCTCCTTCTGTGATCAGATCAACAAGTGGTCTCCAACTTGCAGGTT 352
DB 35072 ACAACAGCAACAGCAACAGCAACAAACAGCAACAGCAACAGCAACAGCAACAA 35013
QY 353 ACACCATCATCGACTCCGTCACGAGCAACTCCGGTTGGTTCATCGGTATCATCTCCA 412
DB 35012 ACAGCAACACACACAGCAGCAGCAACATACACAGCAACACACACACATCAGCA 34953
QY 413 TCCTGTCTTCCACCTTGAGAGCAGAGGAGCTCCGAGAGTCCATCAACTTCTCCTTACG 472
DB 34952 ACAACACAGCAACACACAGCAACAAACAGCAACAGCAACAGCAACAGCAACAA 34893
QY 473 ACATTCACCAACAGCGTCTCTGTTTACAAAGTGGTTCTTGTCACCGTCCACCAACA 532
DB 34892 ACACAGCAACACACACAGCAACACACAGCAACAGCAACACACACAGCAACACA 34833
QY 533 TGATGGGTAAACATGAGATCTACATCAACGGTAACTGATCGACACCATCAAGGTCAAG 592
DB 34832 GCAGCAACAAACACAGCAACAAACAGCAGCAACAAACAGCAACAGCAACAAACA 34773
QY 593 AGTTGACCGGTATCAACTTCTCAAGACCATCACTTCGAGATCAACAGATCCAGACA 652
DB 34772 ACAGCAACAAACAGCAACAAACAGCAACAGCAACAAACAGCAACAGCAACAGCA 34713
QY 653 CCGGTCTGATCACTCCGACTCCGCAACATCAACATGTTGGTTCGGTCACTTCTACATCT 712
DB 34712 ACAGCAACAAACAGCAACAAACAGCAACAAACAGCAACAGCAACAAACAGCAAC 34653
QY 713 TCGCCAAGGAGTTGGCGGTAGGACATCAACATCCTCTTCACTCTTGCA---GTACA 769
DB 34652 GCAACAAACAGCAGCAGCAACAAACAGCAACAGCAACAGCAACAGCAACAGCAAC 34593
QY 770 CCAACCTCGTCAAGGACTACTGGGGTAAAGCTGAGATACAAACAGGAGTACTACATGG 829
DB 34592 ACAACACAGCAGCAGCAACACACACAGCAACAAACAGCAACAGCAACAGCAACAGCA 34533
QY 830 TCAACATCGACTACTTGACACAGATACATGTAGGCGCAACTCCAGACAGATCTCTTCAACA 889
DB 34532 ACAACACAAACAGCAACAAACAGCAACAAACAGCAACAAACAGCAACAGCAACAGCA 34473
QY 890 CCAGAGGTAAACAAACAGCACTTCAACGAGGGTTACAAGATCATCATCAAGCGTTCAGAG 949

```

Direct Submission  
 Unpublished  
 2 (bases 1 to 163034)  
 Worley,K.C.  
 Direct Submission  
 Submitted (13-NOV-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 163034)  
 Worley,K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 11, 2002 this sequence version replaced gi:17973419.  
 -----  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 -----  
 Project Information  
 Center project name: GHWE  
 Center clone name: CH230-103G20  
 -----  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 79006 bases at least Q40  
 Consensus quality: 84121 bases at least Q30  
 Consensus quality: 88465 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 82 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved  
 \*  
 \* 1 1382: contig of 1382 bp in length  
 \* 1383 1482: gap of unknown length  
 \* 1483 2616: contig of 1134 bp in length  
 \* 2617 2716: gap of unknown length  
 \* 2717 4049: contig of 1333 bp in length  
 \* 4050 4149: gap of unknown length  
 \* 4150 5644: contig of 1495 bp in length  
 \* 5645 5744: gap of unknown length  
 \* 5745 6913: contig of 1169 bp in length  
 \* 6914 7013: gap of unknown length  
 \* 7014 8388: contig of 1375 bp in length  
 \* 8389 8489: gap of unknown length  
 \* 8490 10138: contig of 1650 bp in length  
 \* 10139 10238: gap of unknown length  
 \* 10239 11734: contig of 1496 bp in length  
 \* 11735 11834: gap of unknown length  
 \* 11835 13287: contig of 1453 bp in length  
 \* 13288 13387: gap of unknown length  
 \* 13388 14709: contig of 1322 bp in length  
 \* 14710 14809: gap of unknown length  
 \* 14810 15861: contig of 1052 bp in length  
 \* 15862 15961: gap of unknown length  
 \* 15962 17518: contig of 1557 bp in length  
 \* 17519 17618: gap of unknown length  
 \* 17619 19300: contig of 1682 bp in length  
 \* 19301 19400: gap of unknown length  
 \* 19401 20723: contig of 1329 bp in length  
 \* 20730 20823: gap of unknown length  
 \* 20830 21923: contig of 1100 bp in length  
 \* 21930 22023: gap of unknown length  
 \* 22030 23415: contig of 1386 bp in length  
 \* 23416 23515: gap of unknown length



*	22516	25034:	contig	of 1519	bp in length
*	25035	25134:	gap	of unknown	length
*	25135	26213:	contig	of 1079	bp in length
*	26214	27303:	gap	of unknown	length
*	26314	27705:	contig	of 1392	bp in length
*	27706	27805:	gap	of unknown	length
*	27806	29246:	contig	of 1441	bp in length
*	29247	29346:	gap	of unknown	length
*	29347	30523:	contig	of 1177	bp in length
*	30524	30623:	gap	of unknown	length
*	30624	32166:	contig	of 1543	bp in length
*	32167	32266:	gap	of unknown	length
*	32267	33574:	contig	of 1408	bp in length
*	33575	33774:	gap	of unknown	length
*	33775	35023:	contig	of 1249	bp in length
*	35024	35123:	gap	of unknown	length
*	35124	37010:	contig	of 1887	bp in length
*	37011	37110:	gap	of unknown	length
*	37111	38878:	contig	of 1768	bp in length
*	38879	38578:	gap	of unknown	length
*	38979	40229:	contig	of 1351	bp in length
*	40330	40429:	gap	of unknown	length
*	40430	41896:	contig	of 1467	bp in length
*	41897	41956:	gap	of unknown	length
*	41997	43475:	contig	of 1479	bp in length
*	43476	43575:	gap	of unknown	length
*	43576	44684:	contig	of 1109	bp in length
*	44685	44784:	gap	of unknown	length
*	44785	46337:	contig	of 1853	bp in length
*	46338	46737:	gap	of unknown	length
*	46738	47387:	contig	of 1250	bp in length
*	47388	48087:	gap	of unknown	length
*	48088	49155:	contig	of 1068	bp in length
*	49156	49255:	gap	of unknown	length
*	49256	50589:	contig	of 1334	bp in length
*	50590	50889:	gap	of unknown	length
*	50890	51988:	contig	of 1299	bp in length
*	51989	52088:	gap	of unknown	length
*	52089	53066:	contig	of 1878	bp in length
*	53067	54066:	gap	of unknown	length
*	54067	55332:	contig	of 1266	bp in length
*	55333	55432:	gap	of unknown	length
*	55433	56979:	contig	of 1547	bp in length
*	56980	57079:	gap	of unknown	length
*	57080	58450:	contig	of 1371	bp in length
*	58451	58550:	gap	of unknown	length
*	58551	61372:	contig	of 2722	bp in length
*	61373	61372:	gap	of unknown	length
*	61373	62639:	contig	of 1367	bp in length
*	62640	62739:	gap	of unknown	length
*	62740	64180:	contig	of 1441	bp in length
*	64181	64280:	gap	of unknown	length
*	64281	65566:	contig	of 1286	bp in length
*	65567	65666:	gap	of unknown	length
*	65667	66958:	contig	of 1292	bp in length
*	66959	67058:	gap	of unknown	length
*	67059	68733:	contig	of 1875	bp in length
*	68734	68833:	gap	of unknown	length
*	68834	70507:	contig	of 1674	bp in length
*	70508	70607:	gap	of unknown	length
*	70608	71880:	contig	of 1273	bp in length
*	71881	71980:	gap	of unknown	length
*	71981	73275:	contig	of 1295	bp in length
*	73276	73375:	gap	of unknown	length
*	73376	74950:	contig	of 1575	bp in length
*	74951	75050:	gap	of unknown	length
*	75051	76776:	contig	of 1726	bp in length
*	76777	76876:	gap	of unknown	length
*	76877	77949:	contig	of 1073	bp in length
*	77950	78049:	gap	of unknown	length
*	78050	79783:	contig	of 1734	bp in length
*	79784	79883:	gap	of unknown	length
*	79884	82083:	contig	of 2200	bp in length

ACCESSION AC096212.4 GI-21723349  
 VERSION HIG; HIGS\_PHASE1.  
 KEYWORDS Norway rat.  
 SOURCE Rattus norvegicus  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 165337)  
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-oshman, F.R., Allen, C., Ausbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T., Barabara, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burrell, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Hernandez, O., Hodgson, A., Hoques, M., Hollaway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, J., King, L., Korvah, J., Kovar, C., Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, F., Massey, E., Mathiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojucan, I., Rolfe, M., Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaiker, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalob, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 165337)  
 AUTHOR WORLEY, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 165337)  
 AUTHOR WORLEY, K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jul 10, 2002 this sequence version replaced gi:17943903.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: 3ERC  
 Center clone name: CH230-23N22  
 ----- Summary Statistics

Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 96557 bases at least Q40  
 Consensus quality: 103304 bases at least Q30  
 Consensus quality: 107916 bases at least Q20  
 -----  
 NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 68 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1019: contig of 1019 bp in length  
 \* 1020 1119: gap of unknown length  
 \* 1120 2182: contig of 1063 bp in length  
 \* 2183 2282: gap of unknown length  
 \* 2283 3490: contig of 1208 bp in length  
 \* 3491 3590: gap of unknown length  
 \* 3591 4592: contig of 1002 bp in length  
 \* 4593 4592: gap of unknown length  
 \* 4593 5977: contig of 1285 bp in length  
 \* 5978 6077: gap of unknown length  
 \* 6078 7186: contig of 1109 bp in length  
 \* 7187 7286: gap of unknown length  
 \* 7287 8414: contig of 1127 bp in length  
 \* 8414 9632: contig of 1169 bp in length  
 \* 9633 9782: gap of unknown length  
 \* 9783 10917: contig of 1135 bp in length  
 \* 10918 12051: contig of 1034 bp in length  
 \* 12052 12151: gap of unknown length  
 \* 12152 13374: contig of 1223 bp in length  
 \* 13375 13474: gap of unknown length  
 \* 13475 14588: contig of 1114 bp in length  
 \* 14589 14688: gap of unknown length  
 \* 14689 16062: contig of 1374 bp in length  
 \* 16063 16162: gap of unknown length  
 \* 16163 17290: contig of 1128 bp in length  
 \* 17291 18879: contig of 1489 bp in length  
 \* 18880 20557: contig of unknown length  
 \* 20558 20657: gap of unknown length  
 \* 20658 21994: contig of 1337 bp in length  
 \* 21995 22094: gap of unknown length  
 \* 22095 23618: contig of 1524 bp in length  
 \* 23619 23718: gap of unknown length  
 \* 23719 24845: contig of 1127 bp in length  
 \* 24846 24945: gap of unknown length  
 \* 24946 26433: contig of 1488 bp in length  
 \* 26434 26533: gap of unknown length  
 \* 26534 28392: contig of 1859 bp in length  
 \* 28393 28492: gap of unknown length  
 \* 28493 30382: contig of 1890 bp in length  
 \* 30383 30482: gap of unknown length  
 \* 30483 32223: contig of 1741 bp in length  
 \* 32224 32323: gap of unknown length  
 \* 32324 33841: contig of 1518 bp in length  
 \* 33842 33941: gap of unknown length  
 \* 33942 35473: contig of 1532 bp in length  
 \* 35474 35573: gap of unknown length  
 \* 35574 36870: contig of 1297 bp in length  
 \* 36871 36970: gap of unknown length  
 \* 36971 38600: contig of 1630 bp in length  
 \* 38601 38700: gap of unknown length  
 \* 38701 40028: contig of 1328 bp in length  
 \* 40029 40128: gap of unknown length



```

Db 188 AAATCAACATCGGTCTAAAGTTAACTTCGATCGGACAGAATCAGATCCAGCTG- 246
QY 221 GTTCTCCGGTGAGGACAGAGGTAAAGTTCATCGTACCCAGAGAACAGATCGTCTACA 280
Db 247 -----TTCAATCTGGAAATCTCCAAAATCGAAGTTATCCTGAAGAATGCTATCGTATACA 301
QY 281 ACTCCATGTAGAGTCCCTTCCTCCATCTCCTCTCGGATCAGAAATCAACAAGTGGTCTCCA 340
Db 302 ACTCTATGTAGAAAATCTCCACCCTCTCTGGATCCGATATCCCGAATACTTCAACT 361
QY 341 ACTTGCC-----AGGTTACACCATCATCGACTCCGTCACGAACAACATCCGGTTGGT 391
Db 362 CCATCTCTCTGAACAATGAATACACCATCATCAACTGCATGGAAAACAATCTGGTTGGA 421
QY 392 CCATCGGTATCATCTCCAACHTCTCTGCTCTCACCCTGAGCAGACGAGGACTCCGAGC 451
Db 422 AAGTATCTCTGAACATACCGGTGAATCATCTGGAATCTGCAAGGACACTCAGGAAATCAAAC 481
QY 452 AGTCATCAACTTCTCTPACGACATCTCCAAACAGGTCCTGGTTTACAACAA---GTGGT 508
Db 482 ACGGTGTTGTTATTCAAATACTCTCAGATGATCAACATCTCTGACTACATCAATCGCTGA 541
QY 509 TCTTCGTACCGGTACCAACAACATGATGGGTAAACATGATGATGATGATGATGATGATGATG 568
Db 542 TCTTCGTTTACCATCACCAACAATCGTCTGATATACTCCAAAATCTACATCAACGCCGCTC 601
QY 569 TGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAAGACCATCACT 628
Db 602 TGATCGACCAAGAACCGATCTCCATCTGGGTACATCCACGCTTCTTAATACATCATGT 661
QY 629 TCGAGATCAACAAGATCCCAAGACCGGTCTGATCACTCCGACTCCGACCAACATCAACA 688
Db 662 TCAAACTGGACGGTTGTGGTGACATC-----ACCGGTACA 697
QY 689 TGTGGATCGGTGACTTCTACATCTTCGCCAAGGAGTTGGAGGTAAAGACATCAACATCC 748
Db 698 TGTGGATCAAAATACTTCAATCTGTTCGCAAGAAGACTGACGAAAAGAATCAAGACC 757
QY 749 TGTTCAACTCCTTTCAGTACACCAACGTCGTCAGGACTACTGGGGTAAAGACTGAGAT 808
Db 758 TGTACGACAACCAAGTCCCAATCTGGTATCTCTGAAAGACTTCTGGGGTGAATGCTGAGT 817
QY 809 ACAACAAGGAGTACTACATGGTCAACATCGACTACTTGAACAGATACATGATCGCCAAC 867
Db 818 ACGACAACCGTACTACATGCTGAATCTGTACGATCCGACAACAAATACGTTGACGTCAC 876

```

Search completed: November 7, 2002, 13:24:47  
Job time : 3520 secs





```

infection; ds.
XX OS Synthetic.
XX OS Clostridium botulinum.
XX
XX Key Location/Qualifiers
XX CDS 10..1362
XX /*tag= a
XX /product= H_C peptide fragment
XX
XX WO200067700-A2.
XX
XX 16-NOV-2000.
XX
XX 12-MAY-2000; 2000WO-US12890.
XX
XX 12-MAY-1999; 99US-0133865.
XX 12-MAY-1999; 99US-0133866.
XX 12-MAY-1999; 99US-0133867.
XX 12-MAY-1999; 99US-0133868.
XX 12-MAY-1999; 99US-0133869.
XX 12-MAY-1999; 99US-0133873.
XX 29-JUL-1999; 99US-0146192.
XX
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
XX Smith JA, Byrne MP, Middlebrook JL, Iapenotiere H;
XX WPI; 2001-016048/02.
XX P-PSDB; AAB04092.
XX
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
XX the heavy chain of botulinum neurotoxin of serotype A-G, useful as
XX vaccine against botulism
XX
XX Claim 2; Fig 5a; 73pp; English.
XX
XX Botulism neurotoxins are translated as a single 150 kDa polypeptide
XX chain and then posttranslationally nicked, forming a dichain
XX consisting of a 100 kDa heavy chain and a 50 kDa light chain which
XX remain linked by a disulfide bond. Nucleic acids encoding the
XX carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
XX chain of botulinum neurotoxin (BoNT) can be used in recombinant
XX expression vectors and expressed in transformed cells to produce
XX peptide antigens useful for eliciting an immune response to give
XX protective immunity against botulinum neurotoxin, which causes
XX botulism. The nucleic acids are expressible in a recombinant
XX organisms such as Escherichia coli or Pichia pastoris. The use
XX of recombinant nucleic acids are advantageous since it eliminates
XX the need to culture large quantities of hazardous toxin-producing
XX bacterium. Production yield from the genetically engineered product
XX is also high and cost of production is lower. The nucleic acids can
XX be derived from Clostridium botulinum serotypes A-G.
XX
XX Sequence 1371 BP; 397 A; 416 C; 265 G; 293 T; 0 other;
XX
XX Query Match 100.0%; Score 1371; DB 22; Length 1371;
XX Best Local Similarity 100.0%; Pred. No. 4.1e-289;
XX Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps
XX
XX QY 1 GAATTCAGATGACCATCCCATTCACATCTTCTCCTACACCAACAACCTCCCTGTTGAAG 60
XX |||||||
XX Db 1 GAATTCAGATGACCATCCCATTCACATCTTCTCCTACACCAACAACCTCCCTGTTGAAG 60
XX |||||||
XX QY 61 GACATCATCAACGAGTAGTACTTCAACAACATCAACGACTCCCAAGATCTGTCCCTGCAGAAC 120
XX |||||||
XX Db 61 GACATCATCAACGAGTAGTACTTCAACAACATCAACGACTCCCAAGATCTGTCCCTGCAGAAC 120
XX |||||||
XX QY 121 CGTAGACACACTTGGTTCGACACCTCCGTTACACGCCGAGGTCTCCGAGGAGGGTGAC 180
XX |||||||
XX Db 121 CGTAGACACACTTGGTTCGACACCTCCGTTACACGCCGAGGTCTCCGAGGAGGGTGAC 180
XX |||||||
XX QY 181 GTCCAGCTGAACCAAACTTCCCATTCGACTTCAAGCTGGGTTCTCCTCCGTTGAGGACAGA 240

```

[illegible]

Db 1261 TACAGACACAACTACTTGGTTCCAACTGTCAAGCAGGGTAACATACGCCCTCCCTGGTGGAG 1320  
 QY 1321 TCCACTCCACCCACTGGGATTGCTCCCACTCTCCAGAGTAATAGGAATTC 1371  
 Db 1321 TCCACTCCACCCACTGGGATTGCTCCCACTCTCCAGAGTAATAGGAATTC 1371  
 RESULT 3  
 AAX25521  
 ID AAX25521 standard; DNA; 3950 BP.  
 XX AAX25521;  
 AC AAX25521;  
 XX 02-AUG-1999 (first entry)  
 DF DNA coding for modified botulinum toxin rBont/C.  
 DE Botulinum toxin; botulism; rBont/C; vaccine; drug delivery;  
 XX mutant; ss.  
 KW Clostridium botulinum.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FT CDS 32..3907  
 FT /\*tag= a  
 XX WO920306-A1.  
 XX 29-APR-1999.  
 PD 16-OCT-1998; 98WO-US21897.  
 XX 20-OCT-1997; 97US-0954302.  
 PR (UYJE-) UNIV JEFFERSON THOMAS.  
 XX Kiyatkin N, Maksymowych A, Simpson L;  
 PI WPI; 1999-302646/25.  
 DR P-PSDB; AAY05814.  
 XX Modified toxin useful for systemic delivery of oral vaccines and  
 PT therapeutic agents  
 XX Example 1; Page 26-31; 37pp; English.  
 PS This DNA sequence codes for a modified serotype C botulinum toxin,  
 CC termed rBont/C (see AAY05814), in which amino acids His-229, Glu-230  
 CC and His-233 of the native sequence are substituted by Gly, Thr and  
 CC Asn, respectively, i.e. the zinc binding motif (see AAY05817) of  
 CC the light chain holotoxin is modified, resulting in loss of  
 CC endoprotease activity. DNA coding for the modified botulinum toxin  
 CC was assembled from 3 separate toxin fragments using PCR and  
 CC site-directed mutagenesis. The modified recombinant botulinum  
 CC toxin maintains its ability to translocate from the gut into the  
 CC general circulation but is non-toxic. It can be used as an oral  
 CC vaccine for antigenic peptides including botulinum toxin (i.e. an  
 CC oral vaccine for botulism) or for the oral delivery of other  
 CC therapeutic agents to the general circulation.  
 XX Sequence 3950 BP; 1589 A; 410 C; 616 G; 1335 T; 0 other;  
 SQ  
 Query Match 48.5%; Score 665.2; DB 20; Length 3950;  
 Best Local Similarity 68.3%; Pred. No. 3e-135;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
 QY 13 ACCATCCCATTCACATCTCTCTACACCAACTCCCTGTTGAAGGACATCATCAAC 72  
 Db 2558 ACRATACCCCTTTAATATTTTTCATATACTAATAATCTTTTATTAAGAGATATAATTAAT 2617  
 QY 73 GAGTACTTCAACACATCAACAGACTCCAGACTCCCTCCACCAACCGTAAGACACC 132

Db 2618 GAATATTTCAATTAATTAATGATTCAAAAAATTTTGGCCCTACAAAAACAGAAAAAATACT 2677  
 QY 133 TTGCTCGACACCTCCCGGTTAAACGCCGAGGTCTCCGAGAGGGTGACCTCCAGCTGAC 192  
 Db 2678 TTAGTGGATACATCAGGATATAATGCAAGAGTGAAGAGGGGATGTTTCACTTTAAT 2737  
 QY 193 CCAATCTCCCATTCGACTTCAAGCTGGGTCTCCCGGTGAGGACAGAGGTAAAGTCAATC 252  
 Db 2738 CCAATATTTCCATTTGACTTTAAATAGGTAGTTTCAGGGGAGGATAGAGTAAGTTATA 2797  
 QY 253 GTCACCCAGAACGAGAACATCGTCTCAACATCCATGTCAGTGCCTTCCCATCTCCCTTC 312  
 Db 2798 GTAACCCAGAAATGAAAAATATTGTATATAATCTATGTAAGATTTTATAGCATTTAGTTT 2857  
 QY 313 TGGATCAGAACTCAACAGTGGTCTCCCACTTCCAGGTTCACACCATCATCGACTCCGTC 372  
 Db 2858 TGGATTAGATAATAATAATGGTAAGTAATTTACCTGGATATATCTATTAATGATAGTGT 2917  
 QY 373 AAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCTGGTCTTCCACCTCGAG 432  
 Db 2918 AAAAATAACTCAGGTGGAGTATAGGTATTATTAGTAATTTTATTAGTATTACTTTTAAAA 2977  
 QY 433 CAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCCTACGACATCTCCCAACAACGCTCCT 492  
 Db 2978 CAAAATGAAGATAGTGAACAAAGTATAAATTTTAGTTATGATATATCAATAATGCTCCT 3037  
 QY 493 GGTACAAACAAGTGTCTTCTCGTCCACCGTCAACCAACAATGATGGGTACATGAAGATC 552  
 Db 3038 GGATACAATAAATGGTTTTTGTAACTGTACTAACAAATATGATGGGAAATATGAAGATT 3097  
 QY 553 TACATCAACGGTAACTGATCGACACCATCAAGTCAAGGAGTTGACCGGTATCAACTTC 612  
 Db 3098 TATATAAATGGAAAAATTAATAGATACATATAAAGTTAAAGTAACACTGGAATTAATTTT 3157  
 QY 613 TCCAAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCTGTATCACTCCGAC 672  
 Db 3158 AGCAAACTATAACATTTGAATAAATAAATCCAGATACCGGTTTGTACTTACTTCAGAT 3217  
 QY 673 TCCGACACATCAACAATGTCGATCCGTGACTTCTACATCTTCGCCAGAGGAGTTGGACGGT 732  
 Db 3218 TCTGATACATCAATATGTGGATAGAGATTTTATATATTTGCTAAAGAAATAGATGGT 3277  
 QY 733 AAGGACATCAACATCTCTTCAACTCTCTGAGTACACCAACGCTCTCAAGGACTACTGG 792  
 Db 3278 AAAGATATTAATATTTTAAATAGCTTGCAATATACTAATGTTGTAAAGATTTATTGG 3337  
 QY 793 GGTAACGACCTGAGATACAACAAGAGTACTACATGTTCACTACACTGACTTCTGAACAGA 852  
 Db 3338 GGAAATGATTTAAGATATAATAAAGAAATATTATGTTAATATAGATTATTAAATAGA 3397  
 QY 853 TACATGTAGCCCACTCCAGACAGATCTCTTCAACACACAGACGTAAACAACGACTTC 912  
 Db 3398 TATATGTATGCGAATCAGACAAATTTCTTTTAAATACAGTAGAATAATAATGACTTC 3457  
 QY 913 AACGAGGTTCAAGATCATCATCAAGCGTATCAGAGTACACCAACGACACGACAGTC 972  
 Db 3458 AATGAGGATATAAATAATTAATAAAGAAATCAGAGAAATCAAAATGATACATAGAGTA 3517  
 QY 973 AGAGGTGGTACATCTCTGACTTCGACATGACTATCAACAACAAGGCTTCAACCTGTC 1032  
 Db 3518 CGAGGAGGAGATTTTATATTTTGTATGACAAATTAATAACAAGCATATAATTTGTTT 3577  
 QY 1033 ATGAAGAAGAGACCATGTACGCCGACAAACCTCCACGAGGACATCTAGCCCATCGGT 1092  
 Db 3578 ATGAAGAATGAACATCTGTATGACATATCATAGTACTGAAGATATATATGCTATAGGT 3637  
 QY 1093 CTGCGTGACGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCCAATGAAC 1152  
 Db 3638 TTAAGAGAACAAACAAAGGATATAATGATATATATTTCAATACACCAATGAAT 3697  
 QY 1153 AACACTTACTACTACGCTTCCAGATCTTCAAGTCCCACTTCAACGGGTGAGAACATCTCC 1212  
 Db 3698 AATACITATTATTACGCATCTCAAAATTTAAATTTAAATTTTAAATTTTAAATTTTCT 3757





Db 1195 ATCATAGTACTGAGATATATATGCTATAGGTTTAAGAGACAACAACAGGATATAAATG 1254  
Qy 1121 ACAACATCATCTCCAGATCCAGCCAAATGAACAACACTTACTACTAGCTTCCCAAGATCT 1180  
Db 1255 ATAATATTATATTCAATACAAACCAATGAATAATATTATTATACGCATCTCAATAT 1314  
Qy 1181 TCAAGTCCAACTTCAAGCGGTGAGAACATCTCCGGTATCTGTTCCATCGGTACTACAGAT 1240  
Db 1315 TTAATCAATTTTAAATGGAGAAATATTTCTGGAAATATGTTCAATAGGTACTTATCGTT 1374  
Qy 1241 TCCGTCTGGGTGGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1300  
Db 1375 TTAGACTTGGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1434  
Qy 1301 ACTAGCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1360  
Db 1435 ATTATGCTCTAT 1494  
Qy 1361 AATAG 1365  
Db 1495 AAAAG 1499

## RESULT 5

AAZ98630

ID AAZ98630 standard; DNA; 1208 BP.

XX AC

AAZ98630;

XX AC

20-JUN-2000 (first entry)

XX DT

XX DE C. botulinum type D toxin nucleotide sequence.

XX KW

XX Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;

XX OS

XX Clostridium botulinum.

XX PN

XX WO200005252-A1.

XX PD

XX 03-FEB-2000.

XX PF

XX 20-JUL-1999; 99WO-IB01301.

XX PR

XX 22-JUL-1998; 98ZA-0006538.

XX XX

XX (AGRI-) AGRIC RES COUNCIL.

XX PA

XX De Bruyn EE, Botha AD;

XX PI

XX WPI: 2000-205375/18.

XX DR

XX P-PSDB; AAY78982.

XX XX

XX Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin

XX PT

XX type D, useful in vaccines for protection against botulism, comprises

XX PT

XX at least one amino acid mutation not present in the wild type D

XX PT

XX neurotoxins

XX PS

XX Claim 8; Page 54-57; 66pp; English.

XX XX

XX This sequence represents a nucleotide sequence of a synthetic gene

XX CC

XX encoding a non-toxic immunogenic derivative of Clostridium botulinum type

XX CC

XX D toxin (BoNT). Botulinum neurotoxin causes botulism poisoning in cattle

XX CC

XX and sheep, and usually results in the death of the affected or poisoned

XX CC

XX animal. The non-toxic immunogenic fragments of the C. botulinum

XX CC

XX neurotoxin are useful in vaccines to protect animals (e.g. humans,

XX CC

XX cattle, sheep, pigs) against BoNT type D poisoning. The non-toxic

XX CC

XX fragments can be produced relatively simply and inexpensively

XX CC

XX (specifically by fermentation techniques). As the fragments are not

XX CC

XX toxic the risk to production staff is reduced.

XX XX

XX Sequence 1208 BP; 416 A; 203 C; 216 G; 373 T; 0 other;

Query Match 38.4%; Score 526.4; DB 21; Length 1208;  
Best Local Similarity 65.9%; Pred. No. 3.7e-105;  
Matches 799; Conservative 0; Mismatches 401; Indels 12; Gaps 2;  
Qy 157 GCGAGGTCTCCGAGGAGGTGACGTCCAGCTGAACCCAACTTCCCAATTCGACTTCAAG 216  
Db 4 GCAGAAAGTTAGAGTTTCGGGAAACGTGCACTTGAATCCCTATCTTCCATTTGACTTTAAG 63  
Qy 217 CTGGTTCTCCCGGTGAGGACAGAGTAAGGTTCATCGTCACOCAGAGAGAGACATCGTC 276  
Db 64 TTGGGTAGCTCTGCTGATGACCGTGCACAAATTAATGTAACCTCAAAACGAAATATGTT 123  
Qy 277 TACAACCTCCATGACAGTCCCTTCCATCTCCTCTCTGATCAGATCAACAGTGGGTC 336  
Db 124 TATAATGCTATGATGAGTCTTCTCTATAGCTTCTGATCAGATTAATAATATGAGTT 183  
Qy 337 TCAACCTGCGAGGTTCACCATCATCGCTCCGTCACAGAACAACTCCGGTTGGTCCATC 396  
Db 184 TCAATTTACAGGTTATACCATCATCGACGGTTAAATAATTCAGGATGGTCTATT 243  
Qy 397 GGTATCATCTCCAACTTCCCTGGTCTTCCCTGACGACAGACGAGGACTCCGAGGATCC 456  
Db 244 GGCATAATTAGCAATTTTCTAGTTTTTACACTAAAGCAAAAGAAATAGCGAGGAGAC 303  
Qy 457 ATCAACTTCTCTACGACATCTCCAAACGCTCTCTGGTTACAAAGTGGTTCTTCGTC 516  
Db 304 ATTAATTTTCTTATGACATTTCAAGAGCGCTGCTGGTTATAATAATGTTCTTTGTA 363  
Qy 517 ACCGTCAACCAACATGATGGGTATCATGATGATGATGATGATGATGATGATGATGATG 576  
Db 364 ACCATTACCAACATGATGGGTATGATGATGATGATGATGATGATGATGATGATGATG 423  
Qy 577 ACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAAAGACCATCACTTCGAGATC 636  
Db 424 ACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAAAGACCATCACTTCGAGATG 483  
Qy 637 AACAGATPCCAGACACCGGTCTGATCACCTCCGACTCCGACACATCAACATGTCGATC 696  
Db 484 AATAAAATCCCGAACACAGGATTAATAACGTCTGATCTGACAACTCAATATGTGGATT 543  
Qy 697 CGTGACTTCTACATCTTCGCCAGGAGTTGACGGTAAAGGACATCAACATCCCTGTTCAAC 756  
Db 544 AGAGATTCTATATTTTCGCAAGAAATTAGATGACNAGGACATTAATATCCCTTTCAAT 603  
Qy 757 TCCCTTGAGTACACCAACGTGTCAGGAGTACTGGGGTAAACACCTGAGATCACACAAG 816  
Db 604 TCACITCAGTACACTAATGTTGTTAAGGATTAATGGGTAATGATTTAAGATATGATAAG 663  
Qy 817 GAGTACTACATGGTCAACATCGACTACTTGAACAGATACATGTAGCGCAACTCCAGACAG 876  
Db 664 GAGTACTACATGATTAATGTAATTAATGATGATGATGATGATGATGATGATGATGATG 723  
Qy 877 ATCGTCTTCAACACAGACGTAAACAACAGACTTCAACGAGGTTTACAAGATCATCATC 936  
Db 724 ATCGTCTTCAACACTCGTAAATAATAACAACGATTTCAACGAGGTTTACAATAATCATTA 783  
Qy 937 AAGCGTATCAGAGGTAAACAACAGACAGACAGTCAAGAGTGTGTGATCCTGTACTTC 996  
Db 784 AAAGAATCCCGTGGAAATACAAATGATGATGATGATGATGATGATGATGATGATGATG 843  
Qy 997 GACATGACTATCAACAACAGCGCTTCAACCTGTTCATAGAGACGAGACCATGTAGGCC 1056  
Db 844 AATAACAAGATAGATAATAAGCAGTACTCTTTAGGCATGTATAAACCGTCAA----- 895  
Qy 1057 GACAACACTCCCGAGGACATCTACCCCTCGTCTGCGTGGAGCAGACCAAGGACATC 1116  
Db 896 -GAAACTTAGGACAGACCTGTTTCCATGGGTGCAATAGATCAACCGATGGAGCAAT 954  
Qy 1117 AACGCAACATCATCTTCCAGATCCAGCAATGAACACACTTACTACTACGTTCCAG 1176  
Db 955 CGTAATATGGATCATTCATTTATCAACCATGTAATACGTTCCGATTACTACGCGAGCAA 1014  
Qy 1177 ATCTTCAAGTCAACTTCAACGGTGAGAACATCTCCGGTATCTCTCCATCGGTACCTAC 1236

```

Db 1015 TTATTTCTTCTAGCAACGCCAGCACATCGTTTGGAAATACATCAATCGGCTCATAT 1074
QY 1237 AGATCCGCTGGGTGGTG---ACTGGTACAGACACAACACTACTTGGTTCCAACTGTCAG 1293
Db 1075 TCCTTTAACTGGGTGATGATTACTGTGTTAAACCATGAATACTTGATCCAGTAATTTAA 1134
QY 1294 CAGGTAACACGCTCTGCTGGAGTCCACTTCCACCCACTGGGATTCGTCGCCAGTC 1353
Db 1135 ATCGAATACGCTTCACTACGTAATCACTTACACATGGGTCTTGTCCAGCG 1194
QY 1354 TCCGAGTAATAG 1365
Db 1195 TCTGAGTAAAG 1206

RESULT 6
AAZ98631
ID: AAZ98631 standard; DNA; 1440 BP.
XX
AC AAZ98631;
XX
XX 20-JUN-2000 (first entry)
XX
DE Type D BoNT non-toxic immunogenic derivative nucleotide sequence.
XX
KW Botulinum neurotoxin type D; BoNT; botulism; non-toxic; vaccine; poison;
KW protect; ds.
XX
OS Clostridium botulinum.
XX
XX W020005252-A1.
XX
XX 03-FEB-2000.
XX
XX 20-JUL-1999; 99WO-1B01301.
XX
XX 22-JUL-1998; 98ZA-0006538.
XX
XX (AGRI-) AGRIC RES COUNCIL.
XX
XX De Bruyn EE, Botha AD;
XX
XX WPI: 2000-205375/18.
XX
XX P-PSDB; AAY78982.
XX
XX Non-toxic immunogenic derivative of Clostridium botulinum neurotoxin
XX type D, useful in vaccines for protection against botulism, comprises
XX at least one amino acid mutation not present in the wild type D
XX neurotoxins
XX
XX Claim 8; Page 59-62; 66pp; English.
XX
XX This sequence represents a nucleotide sequence of a synthetic gene
XX encoding a non-toxic immunogenic derivative of Clostridium botulinum type
XX D toxin (BoNT). The sequence includes portions of a suitable plasmid which
XX are immediately upstream and downstream of the gene. Botulinum neurotoxin
XX causes botulism poisoning in cattle and sheep, and usually results in the
XX death of the affected or poisoned animal. The non-toxic immunogenic
XX fragments of the C. botulinum neurotoxin are useful in vaccines to
XX protect animals (e.g. humans, cattle, sheep, pigs) against BoNT type D
XX poisoning. The non-toxic fragments can be produced relatively simply and
XX inexpensively (specifically by fermentation techniques). As the fragments
XX are not toxic the risk to production staff is reduced.
XX
XX Sequence 1440 BP; 477 A; 262 C; 272 G; 429 T; 0 other;
XX
XX Query Match 38.4%; Score 526.4; DB 21; Length 1440;
XX Best Local Similarity 65.9%; Pred. No. 3.8e-105;
XX Matches 799; Conservative 0; Mismatches 401; Indels 12; Gaps 2;
XX
QY 157 GCCGAGGCTCCGAGGAGGTGACGTCCAGCTGAACCCCAATCTTCCCATTCGACTTCAAG 216

```

```

Db 58 GCAGAAAGTTAGAGTTGAGGAAACAGTGCAGTTGAATCCTATCTTTCCATTGACTTTAAG 117
QY 217 CTGGGTTCTCCGTTGAGGACAGAGGTAAAGTCACTGTCACCACGAGACGAGAACATCGTC 276
Db 118 TTGGGTAGCTCTGGTGATACCGGTGGCAAAATATTGTAACTCAAAACGAAATATTGTT 177
QY 277 TACAACCTCCATGATACGAGTCTCTCCATCTCTCTGATCAGATCAACAGTGGGTC 336
Db 178 TATAATGCTATGATGAGCTTTCTCTATTAGCTTCTGGATCAGAAATTAATAATGGGTT 237
QY 337 TCCAACCTTGGCAGGTTACACCATCATCGACTCCGTCAGAAACAACCTCCGGTTGGTCCATC 396
Db 238 TCAAAATTTACAGGTTATACCATCATCGACAGCGTTAAAAATAATTACAGGATGGTCTATT 297
QY 397 GGTATCATCTCCAACCTTCTGGTCTTCAACCTGAAGCAGAAACGAGGACTCCGAGCAGTCC 456
Db 298 GGCATAATTAGCAATTTTCTAGTTTTCACATAAGCAAAACGAAATAGCAGGAGGAC 357
QY 457 ATCAACCTTCTCCACGACATCTCCAACAACGCTCCTGTTTACAAACAAGTGGTTCCTCGTC 516
Db 358 ATTAATTTTCTTATGACATTTCAAAAGAACGCTGCTGGTTATAATAATGGTCTCTTTGTA 417
QY 517 ACCGTCAACAAACAATGATGGTACATGAAGATCTACATCAACGGTAAGCTGATCGAC 576
Db 418 ACCATTACCAACAACATGATGGGTAATATGATGATACATTAACGGAAAGCTAATCGAC 477
QY 577 ACCATCAAGGTCAAGGAGTTCACCGGTATCAACTTCTCCAAGACCATCACCTTCGAGATC 636
Db 478 ACCATCAAGGTGAAGAAATTCACCGGTATCAACTTCTCAAGACCATTAATATGTTGGATT 537
QY 637 AACAAGATCCGACACCGGTCTGATCACTCGGATCCGACATCCGACAAATCAATATGTTGGATC 696
Db 538 AATAAAATCCGAAACACAGAGTTAATACGCTCTGATCTGACAAACATCAATATGTTGGATT 597
QY 697 CGTGACTTCTACATCTTCGCCAAGGAGTTGGAGGTGAAGGACATCAACATCTGTTCAAC 756
Db 598 AGAGATTTCTATATTTTCGCAAAAGAAATAGATGACAGGACATTAATATCTTTTCAT 657
QY 757 TCCTTCAGTACACCAACGTCGTCAGGACTACTGGGGTAACGACCTGAGATACAAACAG 816
Db 658 TCACCTTCAGTACACTAATGTTGTTAAGGATTTATGGGTAATGATTAAAGATATGATTAAG 717
QY 817 GAGTACTACATGGTCAACATCGACTACTTGAACAGATACATGACCGCAACTCCAGACAG 876
Db 718 GAGTACTACATGATTAATGTAATTTACATGATCTGTTACATGCTTAAAGGGTAACCGGT 777
QY 877 ATCGTCTTCAACACGAGAGGTAAACAACAGCTTCAACAGGGGTTACAAGATCATCATC 936
Db 778 ATCGTCTTCAACACTCGTAAATAATACAACGATTTCAACGAAAGGTTACAAAATTTATCATA 837
QY 937 AAGCGTATCAGAGGTAAACCAACGACACGACAGTCAGAGTGGTGACATCTCTGTACTTC 996
Db 838 AAAGAAATCCGTTGGAATACAAATGATAGTCTAGTTCGGTGGTGAAGACGTGCTATATTTT 897
QY 997 GACATGACTATCAACAACAGGCTTCAACACCTGTTTCATGAAGAACGAGACCATGATACGCC 1056
Db 898 AATACAACGATAGATATAAAGCAGTACTCTTTAGGCAATGATATAAACCGTCAA----- 949
QY 1057 GACAACCACTCCACGAGGACATCTACGCCATCGGTCTGCGTGAGCAGACCAACGAGACATC 1116
Db 950 -GAAACTTAGGGACAGACCTTGTTCATTTGGGTGATTAGATCAACCCGATGGACGAAAT 1008
QY 1117 AAGCAACAATCATCTTCCAGATCCAGCAATGAACAACACTTACTACTACGCTTCCAG 1176
Db 1009 CGTAATAATGGATCATTCATTATTAACCATGTAATACGTTTCGATTACTACGGAGCCAA 1068
QY 1177 ATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCCGGTATCTGTTCCTCCATCGGTACCTAC 1236
Db 1069 TTATTTCTTTCTAGCAACGCCAGCAATCGTTTAGGAATACTATCAATCGGCTCATAT 1128
QY 1237 AGATTCCGCTCGGTGGTG---ACTGGTACAGACACAACACTTCTGTTTCCCACTGTCAG 1293
Db 1129 TCCTTTAACTGGGTGATGATTACTGGTTTAAACCATGTAATCTTGTATCCAGTAATTTAA 1188

```

QY 1294 CAGGCTAACTAGCGCTCTCTGCTGAGTCCACTCCACCCACAGGGGATTCGCCAGTC 1353  
 Db 1189 ATCGAACATTAGCTTCAATTACTGAATCAACTTCTACACATTGGGCTTTGTCGCCAGCG 1248  
 QY 1354 TCCGAGTAATAG 1365  
 Db 1249 TCTGAGTAAAG 1260

## RESULT 7

AAAS4487

ID AAAS4487 standard; DNA; 1374 BP.

XX

AC

AAAS4487;

XX

DT 11-APR-2001 (first entry)

XX

DE Botulinum toxin heavy chain C-terminal coding sequence (serotype D).

XX

KW Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;

KW recombinant vector; antigen; immune response; vaccine; bacterium;

KW infection; ds.

XX

XX Synthetic.

OS Clostridium botulinum.

XX

PH Key

CDS

FT

FT

FT

FT

PN

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.

XX Sequence 1374 BP; 367 A; 423 C; 274 G; 310 T; 0 other;  
 SQ

Query Match 30.6%; Score 420.2; DB 22; Length 1374;  
 Best Local Similarity 63.0%; Pred. NO. 5.1e-82;  
 Matches 754; Conservative 0; Mismatches 398; Indels 45; Gaps 5;

QY 4 TTCAGGATGACCATCCCATTCACATCTCTCTCTACCAACAACCTCCCTGTTGAAGGAC 63  
 Db 40 TTGCGAACACCATGCCATTCACATCTCTCTCTACCAACAACCTCCCTGTTGAAGGAC 99  
 QY 64 ATCATCAACGAGTACTTCAACAACATCAACGACTCTCAAGATCTGTCCTCGACAGACCT 123  
 Db 100 ATCATCAACGAGTACTTCAACTCCATCAACGACTCTCAAGATCTGTCCTCGACAGAAC 159  
 QY 124 AAGAACACCTTGGTCGACACTCCGGTTACACGCGAGGCTCCCGAGGAGGCGTACGTC 183  
 Db 160 AAGAACGCTTGGTCGACACTCCGGTTACACGCGAGGCTCCCGAGGAGGCGTACGTC 219  
 QY 184 CAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCCCTCGGTGAGGACAGAGT 243  
 Db 220 CAGTTGAACACCATCTACACCAAGCTTCAAGTTGTCCTCTCCGGTGA-----C 270  
 QY 244 AAGTCATCTGTCACCGAGAGAGACATCTCTACAACTCCATGTACGAGTCTCTCTCC 303  
 Db 271 AAGTCATCTGTCACCGAGAGAGACATCTTGTATCTCCGCACTACGAGAGACTCTCT 330  
 QY 304 ATCTCTCTCTGATCAGATCAACAAGTGGGTCTCCAACTTGC---CAGGTTACACCATC 360  
 Db 331 GTCCTCTCTGATCAAGATCTCCAAAGGACTTGACCAACTCCCAACAGAGTACACCATC 390  
 QY 361 ATCGACTCCGTCAAGACACTCCGGTTGGTCCATCGGTATCATCTCCACTCTCTGGTC 420  
 Db 391 ATCAACTCCATCGAGAGAACTCCGGTTGGAAGTTGTATCGGTAACTGATCAAGTAACTCGAG 450  
 QY 421 TTCACCTGTAAGCAGAGAGGACTCCGAGCAGTCCATCAACTCTCTCCACGACATCTCC 480  
 Db 451 TGGATCTTGCAGGAGCTCAACCGTAACTCAAGTCTCTGATCTTCGACTACTCCGAGTCC 510  
 QY 481 AACAAAGCTCTCTGGTTA---CAACAAGTGGTTCTGTCACCGTCAACCAACACATCATG 537  
 Db 511 TTGTCCACACCGGTTACACCAACAGTGGTTCTGTCACCATCAACCAACATCATG 570  
 QY 538 GGTACATGAAGATCTACATCAACGGTAAAGTGTATCGACACCATCAAGTCAAGGAGGTTG 597  
 Db 571 GGTACATGAAGTGTATCAATCAACGGTAAAGTGTATCGACACCATCAAGTCAAGGAGGTTG 630  
 QY 598 ACCGGTATCAACTCTCCAGACCATCACTTCGAGATCAACAGATCCCAAGATCCCAAGCCGT 657  
 Db 631 GACGAGGTCAAGCTGGACAGACCATCTCTCGGTATCGACGAGAGATCGGACATCGA----- 683  
 QY 658 CTGATCACTCCGACTCCGACCAACATCAACATCGGTGATCGGTGACTCTCTACATCTTCGCC 717  
 Db 684 -----CGAGAACAGTGTGTGGATCGGTGACTCTCTACATCTTCCTCC 726  
 QY 718 AAGGAGTTGGACAGGATCAACATCACTCTGTTCAACTCTCTGAGTACACCAACGTC 777  
 Db 727 AAGGAGTGTCAACAGGAGGATCAACATCGTCTACGAGGTCAGATCTCTGAGGAGGTC 786  
 QY 778 GTCAGGAGTACTGGGTAAAGCCTGAGATACACAGGAGTACTACATGTTCAACATC 837  
 Db 787 ATCAAGGAGTACTGGGTAAAGCCTGAGTTCGACCCGAGTACTACATCATCAAGGAC 846  
 QY 838 GACTACTTGAACAGATACATGTAGCCCAACTCCAGACAGATCTGTTTCAACACAGGAGT 897  
 Db 847 AACTACATCGACCGTTATCTGCCCCAGAGTCCCAAGTCTCTGTTGGTCCAGTACCCT 906  
 QY 898 AACAAACAGGATTCACAGGAGGTTACAAAGATCATCATCAAGGATCATCAGGATACACC 957  
 Db 907 GACCGTTCCAAAGTGTACACCGGTAACCGTATACCATCAAGTCCCTCTCCGCAAGAAC 966

(USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.

Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

WPI; 2001-016048/02.

P-PSDB; AAB04093.

New nucleic acids encoding the carboxy- or amino-terminal portions of  
 the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 vaccine against botulism

Claim 2; Fig 6a; 73pp; English.

Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 chain and then posttranslationally nicked, forming a dichain  
 consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 remain linked by a disulfide bond. Nucleic acids encoding the  
 carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 expression vectors and expressed in transformed cells to produce  
 peptide antigens useful for eliciting an immune response to give  
 protective immunity against botulinum neurotoxin, which causes  
 botulism. The nucleic acids are expressible in a recombinant  
 organisms such as *Escherichia coli* or *Pichia pastoris*. The use  
 of recombinant nucleic acids are advantageous since it eliminates  
 the need to culture large quantities of hazardous toxin-producing  
 bacterium. Production yield from the genetically engineered product

Qy 958 AACGACACAGACTCAGAGGTGGTGCATCCCTGCTACTTCGACATGACTACTCAACACAAAG 1017  
 Db 967 CTTTACTCCCGTATCTGACGGTGAACATCACTCTCTGACATGCTGTACAACTCCCGT 1026  
 Qy 1018 GCTTACAACTGTTGATGAAGAACGACCATGTAGCGCGACAA-----CCACTCCACC 1071  
 Db 1027 AAGTACATGATCATCCGTGACACCGACCATCTAGCGCACCGGGTGGTGGTTC 1086  
 Qy 1072 GAGGACATCTACCCCATCGTCTCGTGTAGCAGACCAAGGACATCAACGACACATCATC 1131  
 Db 1087 CAGAATGTCTACGCCCTCGAAGCTGCATCCACCTGGGTAAGTACGGGTATCGGTA 1146  
 Qy 1132 TTCAGATCCAGCAATGAACACACTTACTACTAGCTTCCAGATCTTCAAGTCC 1188  
 Db 1147 TTCTCCATCAAGACATCGTCTCCAGAACAACTAGTGTCTCCAGATCTTCTCTCC 1203  
 RESULT 8  
 ID AA287217 standard; DNA; 1347 BP.  
 AC AA287217;  
 XX 08-MAY-2000 (first entry)  
 XX DNA encoding synthetic BoNT serotype G (BoNTG) Hc fragment.  
 XX Botulinum neurotoxin; heavy chain; BoNT; serotype G;  
 KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;  
 KW VEE; botulism; vaccine; diagnosis; drug screening; ds.  
 XX Clostridium botulinum.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..1347  
 FT /tag= a  
 FT /product= "Synthetic botulinum neurotoxin serotype G  
 FT (BoNTG); heavy chain C-terminal fragment (Hc)"  
 FT /note= "No stop codon given in the specification"  
 XX  
 PN WO200002524-A2.  
 XX  
 PD 20-JAN-2000.  
 XX  
 PF 09-JUL-1999; 99WO-US15570.  
 XX  
 PR 10-JUL-1998; 98US-0092416.  
 PR 12-MAY-1999; 99US-0133870.  
 XX  
 XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.  
 XX  
 PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;  
 XX  
 XX WPI; 2000-160827/14.  
 DR  
 DR P-PSDB; AAY77139.  
 XX  
 PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum  
 PT toxin serotypes A-G, is used for inducing an immune response against  
 PT botulinum -  
 XX  
 XX Disclosure; Page 47-48; 54pp; English.  
 PS  
 XX The invention relates to novel vaccines that induce a protective immune  
 XX response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F  
 CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant  
 CC DNA construct comprising a C-terminal heavy chain fragment (Hc) from BoNT  
 CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT  
 CC serotypes A-G. In preferred embodiments of the invention, the vector is  
 CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of  
 CC this vector results in the production of large amounts of a protein  
 CC encoded by a sequence cloned into the replicon. The constructs are used  
 CC to produce vaccines against botulism. The proteins can also be used as

CC diagnostic tools for the diagnosis of botulism. The transformed host  
 CC cells can be used to analyse the effectiveness of drugs and agents which  
 CC inhibit toxin effects. The vaccine currently used against botulism is  
 CC dangerous and expensive to produce, and contains formalin, which is very  
 CC painful for the recipient. Also, the vaccine is incomplete, in that only  
 CC 5 of the 7 serotypes are represented in the formulation. The novel  
 CC vaccine overcomes these problems, as it is easily purified, and  
 CC available in large quantities. It is also expressed in the lymph nodes  
 CC for a better immune response. Sequences AA287212-287217 represent  
 CC synthetic DNA sequences encoding BoNT Hc fragments used in the present  
 CC invention. These were optimised for codon usage for expression in yeast.  
 XX  
 XX Sequence 1347 BP; 358 A; 442 C; 263 G; 284 T; 0 other;  
 Query Match 15.9%; Score 217.4; DB 21; Length 1347;  
 Best Local Similarity 52.5%; Pred No. 7.6e-38;  
 Matches 631; Conservative 0; Mismatches 516; Indels 54; Gaps 5;  
 Qy 43 ARCAACTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAACATCAAGACTTCCAA 102  
 Db 4 AAGGACACCATCTGATCCAGGTCTTCAACAACATCATCTCCACATCTCTCCAAACGCC 63  
 Qy 103 ATCCTGTCCCTGCAGAACCGTAAGAACACCTTTGGTCGACACCTCGCGTTTACAAGCCGAG 162  
 Db 64 ATCCTGTCCCTGTCCTACCGTGTGTCGTCTGTATCGACTCTCTCGGTTAGGAGCCACC 123  
 Qy 163 GTCTCCGAGGAGGTGACGTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGTGGGT 222  
 Db 124 ATGAACGTGCTGCCGACGTCTCTTCAACGACATCGGTCAACGGTCAGTTCAAGTGAAC 183  
 Qy 223 TCTCCGGTGCAGAGAGGTGAGTCACTGTCACCCAGAACGAGAACATGCTGTACAAAC 282  
 Db 184 AACTCC-----GAGAACTCCAAACATCACCGCCACCAGTCCCAAGTCTGTCTACGAC 237  
 Qy 283 TCCATGTACGAGTCTTCTCCATCTCTTCTGATCAGAAATCAACAAGTGGTGTCTCCAAAC 342  
 Db 238 TCCATGTTCGACAACTTCTCCATCAACTTCTGGTCCGTACCCCAAGTACAAACAAC 297  
 Qy 343 -----TTGCCAGGTTCACCATCATCGACTCCGCTCTCTGTATCAAGAACGACTCC 384  
 Db 298 GACATCCAGACCTACCTGCAGAACGAGTACACCATCATCTCTGTATCAAGAACGACTCC 357  
 Qy 385 GGTGTGTCATCGGTATCATCTCCAACTTCTCTGCTTCCACCTGAAGCAGAACGAGGAC 444  
 Db 358 GGTGTGAAGTCTCCATCAGGGAACCGTATCATCTGGACCCCTGATCGAGTCAACGCC 417  
 Qy 445 TCCGAGCAGTCCATCAACTTCTCTAGACATCTTCAACAACGCTCTCTGGTTA---CAAC 501  
 Db 418 AAGTCCAAAGTCCATCTTCTCGAGTACTCATCAAGGACAAACATCTCCGACTACATCAAC 477  
 Qy 502 AAGTGGTCTTCTGTCACCGTCAACCAACATGATGGTAAACATGAAGATCTACATCAAC 561  
 Db 478 AAGTGGTCTCCATCACCATCACCACCGCTCTGGTGAAGCCACATCTACATCAAC 537  
 Qy 562 GGTAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAAAGACC 621  
 Db 538 GGTTCCTCGAAGAAGTCCGAGAAGATCTCTGAACCTGGACCGTATCAACTCTCCAAACGAC 597  
 Qy 622 ATCACTTCCGAGTCAACAGATATCCGACACCGGTCTGATCACTCCGACTCCGACACAC 681  
 Db 598 ATGACTTCAAG-----CTGATCACTGTACCGACACCCACC 633  
 Qy 682 ATCAACATGTGGATCCGTGACTTCTTACATCTTGGCCAAAGAGTTGGACGGTAAAGCATFC 741  
 Db 634 AAGTTCCTGTGGATCAAGGACTTCAACATCTTGGTCTGTGAGCTGAACGCCACCGAGGTC 693  
 Qy 742 AACATCTGTTCACCTCTTGCAGTACACCAAGTGTCTAAGGACTTACGGGTACAGAC 801  
 Db 694 TCCCTCTGTACTGATGATCCAGTCTCTCAACCAACCCCTGAAGGACTTCTGGGGAACCCA 753  
 Qy 802 CTGAGATACAAACAGGAGTACTACATGGTCAACATCGACTACTTGAACAGATACATGTAC 861  
 Db 754 CTGGTTACGACCCAGTACTACCTGTTCAACCCAGGGTATGCGAAGACATCTACATCAAG 813

QY 862 GCCAACTCCA---GACAGATCGTCTTCAACACGAGAGGTAAACACACGAGCTTCAACAGAG 918  
 DB 814 TACTTCTCCAAAGGCTCCATGGGTGAGACGCGCCCTCGTACCAACTTCAACACGCGCC 873  
 QY 919 GGTTCACAGATCATCAAGCGTATCAGAGGTAAACACACGAGAGGTAAACACGAGAGGT 978  
 DB 874 ATCAACTACAGAGACCTGTACTCGGTCTGGTTTCATCATCAAGAGGCTCCAACTCC 933  
 QY 979 GGTGATCTCTGTACTTGAGATGACTATCAACACACGAGGCTTCAACCTGTTCATGAAG 1038  
 DB 934 CTAATACATCAACACGAGACATCGTCGGTGGGTGACTATCTACCTGAACATCGAC 993  
 QY 1039 AACGAGACATGTACGCGGACCAACCACTTCCACCGAGAGACATCTACGCAACGCTCGCT 1098  
 DB 994 AACATCTCGAGAGTCTTACCGTGTCTAGTCTGTCTGCTCACTCCAGGAGATCCAGACC 1053  
 QY 1099 GACGAGACAGGAGATCAACGACAAATCATCTTCAGATTCAGGATCCAGCAATGACACACT 1158  
 DB 1054 CAGCTGTCTCTGGGCCCCAATCAAGGAGAGCCCTTACCTTTACGACGCTCTCGAGATCAAG 1113  
 QY 1159 TACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGCTGAGACATCTCCGGTATC 1218  
 DB 1114 AGTACTAGGAGAGACCACTTACATGTCTGATCTGTGCGAGAGGACCAAGACC 1173  
 QY 1219 T 1219  
 DB 1174 T 1174  
 RESULT 9  
 ID AAA54491 standard; DNA; 1368 BP.  
 AC AAA54491;  
 DT 11-APR-2001 (first entry)  
 DE Botulinum toxin heavy chain C-terminal coding sequence (serotype G).  
 KW Botulinum toxin; neurotoxin; heavy chain; recombinant expression;  
 KW recombinant vector; antigen; immune response; vaccine; bacterium;  
 KW infection; ds.  
 OS Synthetic.  
 OS Clostridium botulinum.  
 XX Key Location/Qualifiers  
 FT 10...1359  
 CDS /tag= a  
 FT /product= H\_C peptide fragment  
 PN WO200067700-A2.  
 XX 16-NOV-2000.  
 XX 12-MAY-2000; 2000WO-US12890.  
 PF 12-MAY-1999; 99US-0133865.  
 PR 12-MAY-1999; 99US-0133866.  
 PR 12-MAY-1999; 99US-0133867.  
 PR 12-MAY-1999; 99US-0133868.  
 PR 12-MAY-1999; 99US-0133869.  
 PR 12-MAY-1999; 99US-0133873.  
 PR 29-JUL-1999; 99US-0146192.  
 XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.  
 PA Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;  
 PI WPI; 2001-016048/02.  
 DR P-PSDB; AAB04167.  
 XX

PT New nucleic acids encoding the carboxy- or amino-terminal portions of  
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as  
 XX vaccine against botulism  
 FS Claim 2; Fig 10a; 73pp; English.  
 XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide  
 CC chain and then posttranslationally nicked, forming a dichain  
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which  
 CC remain linked by a disulfide bond. Nucleic acids encoding the  
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy  
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant  
 CC expression vectors and expressed in transformed cells to produce  
 CC peptide antigens useful for eliciting an immune response to give  
 CC protective immunity against botulinum neurotoxin, which causes  
 CC botulism. The nucleic acids are expressible in a recombinant  
 CC organisms such as Escherichia coli or Pichia pastoris. The use  
 CC of recombinant nucleic acids are advantageous since it eliminates  
 CC the need to culture large quantities of hazardous toxin-producing  
 CC bacterium. Production yield from the genetically engineered product  
 CC is also high and cost of production is lower. The nucleic acids can  
 CC be derived from Clostridium botulinum serotypes A-G.  
 XX  
 SQ Sequence 1368 BP; 366 A; 445 C; 267 G; 290 T; 0 other;  
 Query Match 15.9%; Score 217.4; DB 22; Length 1368;  
 Best Local Similarity 52.5%; Pred. No. 7.6e-38;  
 Matches 631; Conservative 0; Mismatches 516; Indels 54; Gaps 5;  
 QY 43 AACAACTCCCTGTTGAAGGACATCATCAAGGAGTACTTCAACACATCAACGACTCCAAAG 102  
 DB 13 AAGGACCATCTCTGATCCAGGTCTTCAACACTACATCTCCAAAGATCTCTCCAAAGCC 72  
 QY 103 ATCCGTGTCCTGACAGACCGTGAAGAACACCTTGGTCGACACCTCCGGTTACACAGCCGAG 162  
 DB 73 ATCCGTGTCCTGTCCTACCGGTGGTGGTCTGATCGACTCTCCGGTTACGAGGACCC 132  
 QY 163 GTCTCCGAGGAGGTGAGTCCAGTCCAGTGAACCCATCTTCCATTCCGACTTCAAGCTGGGT 222  
 DB 133 ATGAACGTGCGGTTCGAGGTATCTTCAACGACATCGGTACCGTCAAGTCAAGCTGAAC 192  
 QY 223 TCTCCGTGTGAGGACAGAGTGAAGTATCGTTCACCGAAGACGAGACATCGTCTCAAC 282  
 DB 193 AACTCC-----GAGAACTCCAAACATCACCGCCACCAAGTCCGCTCTACGAC 246  
 QY 283 TCCATGTACGAGTCTCTCTCCATCTCTTGGATCAGATCAACAGTGGGTCTCCAAAC 342  
 DB 247 TCCATGTTCGAACTCTCTCTCCATCTCTTGGGTCCGTACCCCAAGATACACACACAC 306  
 QY 343 -----TTGCCAGGTTACACCATATCGACTCCGCTCAAGAACACACTCC 384  
 DB 307 GACATCCAGACTACCTGCAGAGAGGTACACCATCATCTCTCTGTATCAAGAACGACTCC 366  
 QY 385 GGTGGTCCATCGGTATCATCTCCAACTTCTGGTCTTCCCTGAGGAGAGACGAGGAC 444  
 DB 367 GGTGGAGGTTCCATCAAGGGAACCGTATCATCTGGACCCCTGATCGACGTCAAGCC 426  
 QY 445 TCCGAGCAGTCCATCACTTCTCTAGCATCTCCAAACACGCTCTCTGTTA---CAAC 501  
 DB 427 AAGTCCAAAGTCCATCTTCTCGAGTACTCCATCAAGGACACATCTCCGACTACATCAAC 486  
 QY 502 AAGTGGTCTTCTGTCACCGTCCACCAACACATGATGGGTACATGAGATCTACATCAAC 561  
 DB 487 AAGTGGTCTTCTCATCATCATCAAGGACCGTCTGGGTACGCAACATCTACATCAAC 546  
 QY 562 GGTAAAGTGTGATCGACACCATCAAGGTCAAGGAGTGTGACCGGTATCAACTTCTCCAGACC 621  
 DB 547 GGTTCCTTGAAGAGTCCGAGAGATCTCTGAACCTGGACCGGTATCAACTCTCTCAAGGAC 606  
 QY 622 ATCACTTCGAGATCAACAGATCCAGACACCGGTCTGTATCACTCCGACTCCGACAC 681  
 DB 607 ATCGACTTCAAG-----CTGATCAACTGTACCGGACACCAACC 642

Query Match	14.6%	Score 199.8	DB 22	Length 1317	
Best Local Similarity	56.7%	Pred. No. 5.1e-34			
Matches 497	Conservative	0	Mismatches 332	Indels 48	Gaps
Qy	27	CATCTTCTCCTACACCAACAACCTCCCTGGTTGAAGAGACATCATCAACGAGTAGTACTTCAACAA	86		
Db	6	CAGGATGTCTTACACCAACGACAAGATCCTGTGATCTTTGTTACTTCAACAGCTGTACAAGAA	65		
Qy	87	CATCAACGAGTCCCAAGATCTCTGTCCCTGAGAACCGTATAGACACACTTTGGTCGACACCTC	146		
Db	66	GATCAAGGACAACCTCCATCTTTGGACATGAGATACGAAAAACAATAAGTTTCATCGACATCTC	125		
Qy	147	CGGTTTAAACGCGGAGGTCTCCGAGGAGGTGAGCTCCAGCTGAACCCAAATCTTCCCAATT	206		
Db	126	CGGTTACGGTTCACACATCTCCATCAACGCTGAGCTCATCTACTCCACCAATAGAA	185		
Qy	207	CGACTTCAAGCTGGGTTCCTCCGGTGAGGACAGAGTAGGTCACTGCTCCCGCAGACGA	266		
Db	186	CCAGTTCGGAAATCTACTCCTCC-----AAGCCTCCGAGGTCAACATCGCTCAGAACA	239		
Qy	267	GAACATCGTCTACAACTCCATGTACGAGTCCCTTCCATCTCCCTTCTGGATCAAGATCAA	326		
Db	240	CGACATCATCTACAACGGAAGATACCAAGAACTTCCATCTCCTTCTGGGTCGATATCC	299		
Qy	327	CAAGTGGGTCTCCAACTT-----GCCAGGTTACACCATCATCGACTCCGCT---AA	374		
Db	300	AAAGTACTTTCAACAAGGTCAACCTGAATACAGAGTACACCATCATCGACTGCATCCGTAA	359		
Qy	375	GAACAACCTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTCGTTCCTTCAACCCCTGAAGCA	434		
Db	360	CAATAACTCCGGATGGAAGATCTCCCTGTAATCAACAAGATCATCTGGACCCCTGCAGGA	419		
Qy	435	GAACGAGGACTCCGAGCAGTCCATCAACTTCTCCTACGACATCTCCAACACGCTCCTGG	494		
Db	420	CACCGCGGTGAACAATCAGAAGTTGGTCTTCTCAACTACACCCAGAGTATCTCCATCTCCGA	479		
Qy	495	TTA---CAACAAGTGGTTCTTCGTCACCCGTCACCAACAACATGATGGGTAAACATGAAGAT	551		

```

Db 480 CTACATCAACAAAGTGGATCTTCGTCACCATCAACCAATCAACCGTTTGGGAACCTCCAGAT 539
QY 552 CTACATCAACGGTAAAGTGTGATCGACACCAATCAAGGTAAGGAGTTGACCGGTATCAACTT 611
Db 540 CTACATCAACGGTAAAGTGTGATCGACGAGAGAGTCCATCTCCAAGTCTGGTGACATCCACGT 599
QY 612 CTCGAAGACCATCACTCTCGAGATCAACAAGATCCCGAGACACCGGTCTGTATCACCTCCGA 671
Db 600 CTCGACAAACATTTGTTCAAGATCGTGGTGTAAAGCAGCCCGT----- 646
QY 672 CTCGACAAACATCAACATGTGATCGGTGATCTTACATCTTCCGCAAGGAGTGGACGG 731
Db 647 -----ACGTCGGGATCCGTACTTCAAAAGTCTTCGACACTGAGTTGGTAA 692
QY 732 TAAGGACATCAACATCTCTGTTCAACTCTTCAGTACACCAACAGTGTGTCAGGACTACTG 791
Db 693 GACCGAGATCGAGACCTTGTACTCCGAGGAGCTGACCCATCCATCCTGAAGGACTCTG 752
QY 792 GGGTAACGACCTGAGATACAAACAGGAGTACTATCATGGTCAACATCGACTACTTGAACAG 851
Db 753 GGGTAACCTACCTGCTGTACACAAACGCTTACTACTTCTGTAACCTTGTGGTACCGACAA 812
QY 852 ATACATGTAGCGCAACTCCAGACAGATCGTCTCAAC 888
Db 813 GTCCATCAACCGAGACTCCAACTCTTTGAACATCAAC 849

RESULT 11
AAZ87216
ID AAZ87216 standard; DNA; 1317 BP.
XX
AC AAZ87216;
XX
DT 08-MAY-2000 (first entry)
XX
DE DNA encoding synthetic BoNT serotype F (BoNTF) Hc fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype F;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
XX
OS Clostridium botulinum.
XX
FH Key
FT Location/Qualifiers
FT 10..1308
FT /*tag= a
FT /product= "Synthetic botulinum neurotoxin serotype F
XX (BoNTF) heavy chain C-terminal fragment (Hc)"
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15570.
XX
PR 10-JUL-1998; 98US-0092416.
XX
PR 12-MAY-1999; 99US-0133870.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
XX
DR P-PSDB; AAY7138.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
XX
PS Disclosure; Page 45-46; 54pp; English.
XX

```

```

CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC serotypes A-G. In preferred embodiments of the invention, the vector is
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as
CC diagnostic tools for the diagnosis of botulism. The transformed host
CC cells can be used to analyse the effectiveness of drugs and agents which
CC inhibit toxin effects. The vaccine currently used against botulism is
CC dangerous and expensive to produce, and contains formalin, which is very
CC painful for the recipient. Also, the vaccine is incomplete, in that only
CC 5 of the 7 serotypes are represented in the formulation. The novel
CC vaccine of overcomes these problems, as it is easily purified, and
CC available in large quantities. It is also expressed in the lymph nodes
CC for a better immune response. Sequences AAZ87212-287217 represent
CC synthetic DNA sequences encoding BoNT Hc fragments used in the present
CC invention. These were optimised for codon usage for expression in yeast.
XX
SQ Sequence 1317 BP; 397 A; 375 C; 239 G; 305 T; 1 other;

```

```

Query Match 14.5%; Score 198.2; DB 21; Length 1317;
Best Local Similarity 56.6%; Pred. No. 1.1e-33;
Matches 496; Conservative 0; Mismatches 333; Indels 48; Gaps 5;

```

```

QY 27 CATCTTCTCTACACCAACAACTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAA 86
Db 6 CAGATGTCTTACACCAACGACGACAGATCTTGATCTGTACTTCAACAAAGCTGTACAAGAA 65
QY 87 CATCAACGACTCCAAAGATCTCTGTCGAGAACCGGTAGAACACCTTGGTCGACACCTC 146
Db 66 GATCAAGGACAACTCCATCTTGGACATGAGATACGAAACAAATAAGTTTCATCGACATCTC 125
QY 147 CGGTTACAAACGCGAGGTCTCCGAGGAGGTGAGCTCAGCTGAACCCATCTTCCCAT 206
Db 126 CGGTTACGTTCCAAATCTCCATCAACGCTGAGCTTACATCTATCTCCACCAATAGAAA 185
QY 207 CGACTTCAAGTGGTGGTTCCTCGGTGAGGACAGAGTGAAGTCACTGTCACCCAGAACGA 266
Db 186 CCAGTTCGGAATCTACTCTCTCC-----AAGCCTTCGAGGTCAACATCGCTCAGAACAA 239
QY 267 GAACATCGTCTACAACTCCATGTACGAGTCTCTTCCATCTCTTGTGATCAGATCAAA 326
Db 240 CGACATCATCTACAAACGAGATACCAAGATCTTCTCATCTCTTGGTCTCGTATCCC 299
QY 327 CAAGTGGTCTCCAACTT-----GCCAGGTTACACCATCATCGACTCCGTC---AA 374
Db 300 AAGTACTTCAACAAAGGTCAACCTGAAATAACGAGTACACCATCATCGACTCGATCGTAA 359
QY 375 GAACAACTCCGGTGGTTCCTCATCTCCAACTTCTCTGTTCTTCACTCCGCTGAAGCA 434
Db 360 CAATAACTCCGGATGGAAGATCTCCCTGACTACACAGATCATCTTGGACCTCGAGGA 419
QY 435 GAACGAGGACTCCGAGCAGTCCATCAACTTCTCTAG---GACATCTCCAAACACCTCC 491
Db 420 CACGCGCGGTAAACAATCAGAAGTTGGTCTTCACTACACNAGATGATCTCCATCTCCGA 479
QY 492 TGGTTACAAACAAAGTGGTCTCTCGTCACCGTCACCAACAAACATGATGGTAAACATGAAT 551
Db 480 CTACATCAACAAAGTGGATCTTCTGTCACCATCAACAAATCAACCGTTTGGGAAACCTCCAGAT 539
QY 552 CTACATCAACGTTAAGTGTATCGACACCATCAACGTTCAAGGAGTGTGACCGGTATCAACTT 611
Db 540 CTACATCAACGTTAAGTGTATCGACGAGAGTCCATCTCCAACTTGGGTGACATCCACGT 599
QY 612 CTCGAAGACCATCACTCTCGAGATCAACAAGATCCCGAGACACCGGTCTGTATCACCTCCGA 671
Db 600 CTCGACAAACATTTTGTTCAGATCGTGGTGTAAAGTCTTGTAAAGACACCCCGTT----- 646
QY 672 CTCGACAAACATCAACATGTGGATCGGTCTTACATCTTCCGCAAGGAGTGTGACCG 731

```



```

Db 647 -----ACGTCGGGATCCGTTACTTCAAGAGCTTCGACACGTGAGTTGGTAA 692
Qy 732 TAAGGACATCAACATCTGTTCAACTCCCTTGCAGTACACCAACGCTGCTCAAGGACTACTG 791
Db 693 GACCGAGATCGAGACCTGTACTCCGACGAGCTGACCCATCCATCCTCAAGGACTTCTG 752
Qy 792 GGGTAACGAGCTGAGATACAAACAGGAGTACTACATGCTCAACATCGACTACTTGAACAG 851
Db 753 GGGTAACGAGCTGAGTACAAACAGGAGTACTACTTCTGTAACCTGTTGCGTACCGACAA 812
Qy 852 ATACATGATAGCCCAACTCCACAGACAGATCGCTTTCAC 888
Db 813 GTCCATCAACCCAGACCTCCAACTCTTGAACATCAAC 849

RESULT 12
AAV30576
ID AAV30576 standard; DNA; 1351 BP.
XX
AC AAV30576;
XX
DI 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C gene in pHisBotA(syn).
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulism; ds.
XX
OS Clostridium botulinum serotype A.
OS Synthetic.
XX
FH Key
FT Location/Qualifiers
FT CDS
FT 1..1338
FT /*tag= a
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US15394.
XX
PR 28-AUG-1996; 96US-0704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Thalley BS, Williams JA;
XX
P-PSDB: AAW68391.
XX
WPI: 1998-230234/20.
XX
Host cell containing recombinant expression vector encoding
Clostridium botulinum type B or E toxin - useful to treat humans
and other animals at risk of intoxication with clostridial toxin
Example 29; Page 279-281; 428pp; English.
XX
This is the DNA sequence of the Clostridium botulinum serotype A
toxin C fragment gene contained in plasmid pHisBotA(syn). The
encoded toxin A polypeptide (see AAW68391) has a histidine-tagged
N-terminal extension. The vector was used to express native
(i.e. non-fusion) soluble C fragment in Escherichia coli host
cells. The invention relates to recombinant proteins derived from
C. botulinum toxins. Methods are provided which allow for the
isolation of soluble recombinant proteins free of significant
endotoxin contamination. Preferred hosts for production of
recombinant proteins are E. coli, insect cells and yeast cells.
The recombinant toxins are used as immunogens for the production
of vaccines and antitoxins that are useful in the treatment of
humans and animals at risk of intoxication with clostridial toxin.
XX
Sequence 1351 BP; 407 A; 348 C; 247 G; 349 T; 0 other;

```

```

Query Match 12.4%; Score 169.6; DB 19; Length 1351;
Best Local Similarity 54.4%; Pred. No. 28-27; Indels 42; Gaps 4;
Matches 455; Conservative 0; Mismatches 339;
Qy 44 ACAACTCCCTGTTGAAGGACATCATCAAGAGTACTTCAACACATCAACGACTCCAAGA 103
Db 20 ACATGGCTCCTGCTGTCTACCTTCACTGAATACATCAAGAACATCATCAATCACTCCA 79
Qy 104 TCCGTGCTCCGAGACCGGTAAGAACACCTTGGTCACACCTCCGGTTACACACGGCGAGG 163
Db 80 TCCGTGAACCTGCGCTACGAATCCAATCACTGATCGCTGTCTCGCTACGTTCCAAA 139
Qy 164 TCTCCGAGGAGGCTGAGCTCCAGCTGAACCCCAATCTCCCATTCGACTTCAAGCTGGGT 223
Db 140 TCAACATCGGTTCTAAGTTAACTTCGATCCGATCCAGAAATCAGATCCAGCTG--- 195
Qy 224 CTTCCGGTGAGGACAGGTAAGGTATCGTTCACCAGAACGAGAACATCGTCTCAACT 283
Db 196 --TTCAATCTGGAATCTTCCAAAATCGAAGTTATCTCTGAAATGCTATCGTATACA 253
Qy 284 CCATGTAGAGTCTCTTCCATCTCTTGGATCAGATCAACAGTCAACAGTGGTCTCCA 343
Db 254 CTAATGACGAAACTTCTCCACTCTCTTGGATCGTATCCGAAATFACTTCAACTCCA 313
Qy 344 TGCC-----AGTTACACCATCATCGACTCCGTCAGTCAAGAACACTCCGTTGGTCCA 394
Db 314 TCTCTCTGAAACAATGAATACACCATCATCACTGATGGAAGAAATCTGTTGGTGAAG 373
Qy 395 TCGGTATCATCTCCAACCTCTCTGCTTCACTGAGGAGGAGGAGGAGGAGGAGGAGGAG 454
Db 374 TATCTCTGAACTACGGTGAATCATCTGGACTCTGAGGACACTCAGGAAATCAACAGC 433
Qy 455 CCATCAACTTCTCTCGACATCTCCAACAACCTCTGTTTACACAA---GTGGTTCT 511
Db 434 GTGTTGATTCAAATACTCTCAGATGATCAACATCTCTGACTATCATCACTCGTGTCT 493
Qy 512 TCGTCACTGACCAACAACATGATGGTAAATGATGATGATGATGATGATGATGATGATG 571
Db 494 TCGTACCATCACCACAACATCGTCTGAATTAACCTCAAAATCTACATCAACGGCCGCTGA 553
Qy 572 TCGACACCATCAAGGTCAAGGAGTGTACCGGTATCAACTTCTCCAGACCATCACTCGC 631
Db 554 TCGACCAAGAACCGATCTCCAATCTGGTAAACATCCACGCTTCTAATAACATGTTCA 613
Qy 632 AGATCAACAAGATCCACAGACCGGTCTGATCAGCTCCGACTCCGACACATCAACATGT 691
Db 614 AACTGGACGGTGTCTGTCACACTC-----ACCGTCACTCT 649
Qy 692 GGATCCGTGACTTCTACATCTTCGCAAGGAGTGTGACGGTAAAGGACATCAACATCTGT 751
Db 650 GGATCAATACTTCAATCTCTTCGACAAAGAACTGAACGAAAGAAATCAAGACCTGT 709
Qy 752 TCAACTCTTGCAGTACACCAAGTGTGTCAGGACTACTGGGTACGACCTTGAGATACA 811
Db 710 ACGACACCACTCCAAATCTGTTGATCTCTGAAAGACTCTTGGGTGACTTACCTGCA 769
Qy 812 ACAGGAGTACTACATGTTCAACATCGACTTGTGAACAGATACATGTACGCCAAC 867
Db 770 ACAACCGTACTACTCTGTAATCTGTACGATCCGACAAATACGTTGAGGTCAAC 825

RESULT 13
AAAS4483
ID AAAS4483 standard; DNA; 1323 BP.
XX
AC AAAS4483;
XX
DI 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal coding sequence (serotype A).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;

```



```
PR 14-APR-1995; 95US-0422711.
XX (OPHI-) OPHIDIAN PHARM INC.
XX Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
XX Williams JA;
XX WPI: 1996-230603/23.
XX P-PSDB: AAR95009.
XX Fusion proteins comprising non-toxin protein and part of toxin
XX useful to form anti-toxins against Clostridium botulinum type A, and
XX C. difficile type toxins, and to treat C. difficile intoxication,
XX partic. diarrhoea
XX Example 24; Page 340-342; 434pp; English.
XX A nucleotide sequence (AAR29246) present in vector pETHisa encodes
XX the phisBot fusion protein (AAR95009) comprising a polyhistidine
XX affinity tag and fragment C (see also AAR95008) of the Clostridium
XX botulinum type A neurotoxin. The phisBot protein was expressed
XX in Escherichia coli as a soluble protein and was purified by
XX metal chelate affinity chromatography to obtain a product free
XX of endotoxin contamination that may be useful as an immunogen
XX in vaccine compns.
XX Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;
XX SQ

Query Match 12.4%; Score 169.4; DB 17; Length 1402;
Best Local Similarity 54.4%; Pred. No. 2.2e-27;
Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps 4;

QY 41 CCAACAACCTCCTGTTGAAGGACATCATCAAGAGTACTTCAACAACATCAACGACTCCA 100
DB 68 CTAGCATGGCTCGTCTGCTGTCACCTTCACTGAATACATCAAGAACATCATCAATACCT 127
QY 101 AGATCTGTCCTCCGACCAACGGTAAGACACCTTGTGCACACCTCCGGTTACAGCGCG 160
DB 128 CCATCTCGAAGCTCGGTACGAAATCCAAATCAACCTGATCGACCTGTCTCGCTACGGTTCCA 187
QY 161 AGGTCCTCGAGGAGGTGAGTCCAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGG 220
DB 188 AATCAACATCGGTTCTTAAGTTAACTTCGATCGGATCGACAGATCATGATCCAGCTG- 246
QY 221 GTTCCTCGGTGAGGACAGAGTAAGTATATCTGTCACCCAGACAGAACATCGTCTACA 280
DB 247 -----TTCAATCTGGAATCTTCCAAAATGAAGTTATCTCTGGAAGATGCTATCGTATACA 301
QY 281 ACTCATGTAGAGTCTCTTCCATCTCCTTCTGGATCAGAATCAACAAGTGGTCTCCA 340
DB 302 ACTCTATGTAGAAACTTCTCCACCTCTCTTGGATCCGATCCGGAATATCTCACTCACT 361
QY 341 ACTTGCC-----AGGTTACACCATATCATCGACTCCGTTCAAGAACACTCCGGTTGGT 391
DB 362 CCATCTCTGAAACAATGAATACACCATCATCACTCACTGATGCGATGGAACAATCTCTGTTGA 421
QY 392 CCATCGGTATCATCTCCAACCTCTCTGTTCTCAACCTGAGACAGACGAGCTCGGAC 451
DB 422 AAGTATCTCTGAATAGGTGAATATCTCTGAGTCTGACAGGACATCTAGGAATCAAC 481
QY 452 AGTCCATCAACTTCTCTGACATCTCTCAACAACGCTCTCTGTTTACAACA---GTGCT 508
DB 482 AGCGTGTGTATTCAAAATCTCTCAGATGATCAACATCTCTGACTACATCAATCGCTGA 541
QY 509 TCTTCGTCACCGTCAACAACAATGATGGTGAATCATGATGATGATGATGATGATGATGATGAT 568
DB 542 TCTTCGTTTACCATCAACAATGCTGTAATCACTCAAAATCTACATCAACGCGGCTC 601
QY 569 TGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCACTTCTCTCAAGACATCACT 628
DB 602 TGATCGACCAAAACCGATCTCCAATCTGGTGAATCACTCAACGCTTCTTAATACATCATGT 661
QY 629 TCGAGATCAACAAGATCCAGACACCGGTTCTGATCACTCCGACTCCGACATCAACA 688

RESULT 15
AAV30572
ID AAV30572 standard; DNA; 1402 BP.
XX AAV30572;
AC AAV30572;
XX 07-DEC-1998 (first entry)
XX Clostridium botulinum toxin A fragment C gene in phisBot.
XX Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
XX botulism; ds.
XX Clostridium botulinum serotype A.
XX Synthetic.
XX Key Location/Qualifiers
XX CDS 1..1389
XX FT /*tag= a
XX PN WO9808540-A1.
XX XX
XX PD 05-MAR-1998.
XX PF 28-AUG-1997; 97WO-US15394.
XX PR 28-AUG-1996; 96US-0704159.
XX PA (OPHI-) OPHIDIAN PHARM INC.
XX XX
XX PI Thalley BS, Williams JA;
XX XX
XX DR WPI: 1998-230234/20.
XX P-PSDB: AAW68390.
XX Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX Example 24; Page 265-267; 428pp; English.
XX This is the DNA sequence of the Clostridium botulinum serotype A
XX toxin C fragment gene contained in plasmid phisBot. The encoded
XX toxin A polypeptide (see AAW68390) has a histidine-tagged
XX N-terminal extension. The vector was used to express native
XX (i.e. non-fusion) soluble C fragment in Escherichia coli host
XX cells. The invention relates to recombinant proteins derived from
XX C. botulinum toxins. Methods are provided which allow for the
XX isolation of soluble recombinant proteins free of significant
XX endotoxin contamination. Preferred hosts for production of
XX recombinant proteins are E. coli, insect cells and yeast cells.
XX The recombinant toxins are used as immunogens for the production
XX of vaccines and antitoxins that are useful in the treatment of
XX humans and animals at risk of intoxication with clostridial toxin.
XX Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;
XX SQ
```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:20:20 ; Search time 57 Seconds  
(without alignments)  
7376.383 Million cell updates/sec

Title: US-09-910-186A-9  
Sequence: 1 gaattacagatgaccatcccc.....tctccgagtaataagaattc 1371

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	169.4	12.4	1402	1	US-08-480-604A-25
2	169.4	12.4	1402	2	US-08-405-496A-25
3	169.4	12.4	1402	4	US-08-915-136-25
4	168.8	12.3	1330	1	US-08-480-604A-22
5	168.8	12.3	1330	2	US-08-405-496A-22
6	168.8	12.3	1330	4	US-08-915-136-22
7	136	9.9	5163	3	US-08-700-651-1
8	136	9.9	5163	3	US-08-928-361B-4
9	136	9.9	5318	3	US-08-700-651-2
10	136	9.9	5318	3	US-08-928-361B-3
11	115.4	8.4	5511	3	US-08-928-361B-2
12	115.4	8.4	7334	3	US-08-928-361B-1
13	95	6.9	4483	4	US-08-961-527-363
14	93.4	6.8	2004	1	US-08-471-033-18
15	93.4	6.8	2004	2	US-08-471-044-18
16	93.4	6.8	2004	2	US-08-463-483A-18
17	93.4	6.8	2004	2	US-08-471-046A-18
18	93.4	6.8	2004	2	US-08-470-568B-18
19	93.4	6.8	2004	2	US-08-469-334-18
20	93.4	6.8	2004	3	US-09-300-529-18
21	93.4	6.8	2576	1	US-08-471-033-35
22	93.4	6.8	2576	2	US-08-471-044-35
23	93.4	6.8	2576	2	US-08-463-483A-35
24	93.4	6.8	2576	2	US-08-471-046A-35
25	93.4	6.8	2576	2	US-08-470-568B-35
26	93.4	6.8	2576	2	US-08-469-334-35
27	93.4	6.8	2576	3	US-09-300-529-35

28	93.4	6.8	2555	1	US-08-471-033-17
29	93.4	6.8	2555	1	US-08-471-033-26
30	93.4	6.8	2555	2	US-08-471-044-17
31	93.4	6.8	2555	2	US-08-471-044-26
32	93.4	6.8	2555	2	US-08-463-483A-17
33	93.4	6.8	2555	2	US-08-463-483A-26
34	93.4	6.8	2555	2	US-08-471-046A-17
35	93.4	6.8	2555	2	US-08-471-046A-26
36	93.4	6.8	2555	2	US-08-470-568B-17
37	93.4	6.8	2555	2	US-08-470-568B-26
38	93.4	6.8	2555	2	US-08-469-334-17
39	93.4	6.8	2555	2	US-08-469-334-26
40	93.4	6.8	2555	3	US-09-300-529-17
41	93.4	6.8	2555	3	US-09-300-529-26
42	93.4	6.8	4031	1	US-08-471-033-49
43	93.4	6.8	4031	2	US-08-471-044-49
44	93.4	6.8	4031	2	US-08-463-483A-49
45	93.4	6.8	4031	2	US-08-471-046A-49

## ALIGNMENTS

RESULT 1  
US-08-480-604A-25  
Sequence 25, Application US/08480604A  
Patent No. 5736139  
GENERAL INFORMATION:  
APPLICANT: KINK, JOHN A.  
APPLICANT: THALLEY, BRUCE S.  
APPLICANT: PADHYE, NISHA V.  
APPLICANT: FIRCA, JOSEPH R.  
APPLICANT: STAFFORD, DOUGLAS C.  
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,604A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION NUMBER: US 08/422,711  
FILING DATE: 14-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1402 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1386  
US-08-480-604A-25

Query Match 12.4%; Score 169.4; DB 1; Length 1402;  
Best Local Similarity 54.4%; Pred. No. 8e-29;  
Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps 4;

QY 41 CCAACAACCTCCCTGTTGAAGGACATCATCAAGGACTACTTCAACAACATCAACGACTCCA 100  
DB 68 CTAGCATGGCTCGTCTGCTGCTACCTTCACTGATGATCAAGACATCATCAATACCT 127  
QY 101 AGATCCTGTCCTGCGAGACCGTAAGAACACCTTGGTGCACACCTCCGGTTACAAACCG 160  
DB 128 CCATCCTGAACCTGCGCTACGAATCCAATCACTGATCGAAGCTGCTCGCTACGCTTCCA 187  
QY 161 AGGTCTCCGAGGAGGTGAGTCCAGCTGACCCCAATCTTCCATTCGACTTCAAGCTGG 220  
DB 188 AATCAACATCGGTTCTAAGTTACTTGGATCCGATGACGAAGATCAGATCCAGCTG- 246  
QY 221 GTTCCTCCGCTGAGGACAGAGTAAAGTTCATGTCACCCAGACAGAACATCGTCTACA 280  
DB 247 -----TTCATCTGGAATCTTCCAAAATCGAAGTTATCTCTGAAGAATGCTATCGTATA 301  
QY 281 ACTCCATGACGAGTCTCTCCATCTCTCTGATGATGATGATGATGATGATGATGATGAT 340  
DB 302 ACTCATGTACGAAACTTCTCCACTCTCTGATGATGATGATGATGATGATGATGATGAT 361  
QY 341 ACTTGGC-----AGGTACACCATCATCGACTCCGTCGATCAAGAACACATCCCGTGGT 391  
DB 362 CCATCTCTCGAACAATGAATACACCATCATCACTGATGATGATGATGATGATGATGATG 421  
QY 392 CCATCGGTATCATCTCCACTCTCTGCTGCTTCCCTGATGATGATGATGATGATGATGAT 451  
DB 422 AAGTATCTCTGAACTACGCTGAAATCATCTGGAATCTGGAATCTGGAATCTGGAATCT 481  
QY 452 AGTCCATCAACTCTCTCTAGGATCTCCCAACACGCTCTCTGATGATGATGATGATGAT 508  
DB 482 AGCGTGTGTTTCAATTAATCTCTGATGATGATGATGATGATGATGATGATGATGATG 541  
QY 509 TCTTGTGTCACCGTCACCAACATGATGATGATGATGATGATGATGATGATGATGATGAT 568  
DB 542 TCTTGTGTCACCATCACCAACATGATGATGATGATGATGATGATGATGATGATGATG 601  
QY 569 TGATGACCATCAACAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 628  
DB 602 TGATGACCATCAACAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 661  
QY 629 TCGAGATCAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 688  
DB 662 TCAACTGGAGGTTGCTGAGACTC-----ACGCGTACA 697  
QY 689 TGTGGATCCGCTGATCTACATCTTCGCAAGAGTGTGAGGTTGAGGATGATGATGATGAT 748  
DB 698 TCTGGATCAAAATCTTCAATCTGATGATGATGATGATGATGATGATGATGATGATGAT 757  
QY 749 TGTTCACCTCTTTCAGTACCAACAGCTGCTCAAGGATGATGATGATGATGATGATGATG 808  
DB 758 TGTACGACACCAAGTCCAAATCTCTGGTATCTCTGAAAGACTTCTTGGGGTGAATG 817  
QY 809 ACAACAGGAGTACTACATGTTGATGATGATGATGATGATGATGATGATGATGATGAT 867

DB 818 ACCAGAAACCGTACATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 876  
RESULT 2  
US-08-405-496A-25  
Sequence 25: Application US/08405496A  
Patent No. 5919665  
GENERAL INFORMATION:  
APPLICANT: WILLIAMS, JAMES A.  
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
TITLE OF INVENTION: NEUROTOXIN  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL, LLP  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/405.496A  
FILING DATE: 16-MAR-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 25-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01308  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1402 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1386  
US-08-405-496A-25

Query Match 12.4%; Score 169.4; DB 2; Length 1402;  
Best Local Similarity 54.4%; Pred. No. 8e-29;  
Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps 4;

QY 41 CCAACAACCTCCCTGTTGAAGGACATCATCAAGGACTACTTCAACAACATCAACGACTCCA 100  
DB 68 CTAGCATGGCTCGTCTGCTGCTACCTTCACTGATGATCAAGACATCATCAATACCT 127  
QY 101 AGATCCTGTCCTGCGAGACCGTAAGAACACCTTGGTGCACACCTCCGGTTACAAACCG 160  
DB 128 CCATCCTGAACCTGCGCTACGAATCCAATCACTGATCGAAGCTGCTCGCTACGCTTCCA 187  
QY 161 AGGTCTCCGAGGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 220

Db 188 AAATCAACATCGGTCTAAAGTTAACTTCGATCGACAGAATCATGATCCAGCTG- 246  
Qy 221 GTTCTCGGTGAGGACAGAGGTAAAGTTCGTCACCCAGAGAGAAATCATGCTTACA 280  
Db 247 -----TTCAATCTGAATCTTCCAAATCGAAGTTATCTCTGAGATGCTATCATATA 301  
Qy 281 ACTCGATGATAGATCTCTTCCATCTCTTGGATCAGAAATCAACAGTGGGTCTCCA 340  
Db 302 ACTATATGAGAAATCTTCCACCTCTCTGATCCGTAATCCGGAATATCTTCAACT 361  
Qy 341 ACTTGCC-----AGGTTACACCATCATCGACTCCGTCAAGAAACAATCCGGTGGT 391  
Db 362 CCACTCTCTGAACAATGATACACCATCATCACTGATGGAACAAATCTTGTTGA 421  
Qy 392 CCAATCGGTATCATCTCAACTTCTGCTGTTTACCTGGAAGAGAGAGGACTCCGAGC 451  
Db 422 AAGTATCTCTGAATAGGTTGAATCATCTGACTCTGAGGACACTCAGGAATCAAAAC 481  
Qy 452 AGTCCATCAACTTCTCTACGACATCTCCAAACAGGTCTGTTTACAAACAA---GTGGT 508  
Db 482 AGCGTGTGTTAATCAATCTCTCAGATGATCAACATCTCTGACTATCAATCTGTTGA 541  
Qy 509 TCTTGTCACCGTCACCAACAATGATGGTACATGATGATGATGATGATGATGATGATG 568  
Db 542 TCTTGTCACCGTCACCAACAATGATGGTACATGATGATGATGATGATGATGATGATG 601  
Qy 569 TGATCGACACCATCAAGTCAAGGATGACCGGTATCACTTCTCAAGACATCACT 628  
Db 602 TGATCGACACCATCAAGTCAAGGATGACCGGTATCACTTCTCAAGACATCACT 661  
Qy 629 TCGAGATCAACAAGATCCGACACCGGTGATGATCACTCCGACTCCGACACATCAACA 688  
Db 662 TCAAACTGGCGGTGTCGTGACACTC-----ACCGTACA 697  
Qy 689 TGTGATCCGTGACTTCTACATCTTCCCAAGGATGAGGATGAGGATGAGGATGAGGATG 748  
Db 698 TGTGATCAAAATCTCAATCTGTCACAAAGAACTGAAAGAAAGAAAGAAAGAAAGAAC 757  
Qy 749 TGTCAACTCTTTCAGTACACCAACATGCTCAAGGACTACTGGGTAAACGACTGAGAT 808  
Db 758 TGTAGCAACACAGTCCAAATCTGTTATCTGTAAGAGATCTGTTGGTACTACTGCTGAT 817  
Qy 809 ACACAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 867  
Db 818 ACGAAACACCGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 876

RESULT 3

US-08-915-136-25  
; Sequence 25 Application US/08915136  
; Patent No. 6290960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/915,136  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/480,604  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/405,496  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01763  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1402 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1386  
; US-08-915-136-25

Query Match 12.4%; Score 169.4; DB 4; Length 1402;

Best Local Similarity 54.4%; Pred. No. 8e-29;  
Matches 456; Conservative 0; Mismatches 341; Indels 42; Gaps 4;

Qy 41 CCACAACTCCCTGTTGAGGACATCATCAACGAGTACTTCAACACATCAACGACTCCA 100  
Db 68 CTAGCATGGCTCGTCTGTCTTACCTTCACTGAATACATCAAGAACATCATCAATACCT 127  
Qy 101 AGATCCTGTCCTGCGACAGACCGTAAAGAACACCTTGTGCGACACCTCCGGTTACAAGCGG 160  
Db 128 CCATCCTGAACCTGCGTACGATCCAAATCACTGATCGACCTGTCTCGTACGTTCCA 187  
Qy 161 AGGTCTCCGAGGAGGTGACGTCCAGTGAACCCAAATCTTCCATTTCGACTTCAAGCTGG 220  
Db 188 AAATCAACATCGGTCTTAAAGTTAACTTCGATCCGATCGACAGAAATCAGATCCAGCTG- 246  
Qy 221 GTTCTCGGTGAGGACAGAGTAAAGTATCTGTCACCCAGAGAGAGACATCGTCTACA 280  
Db 247 -----TTCAATCTGGAATCTTCCAAATCGAAGTTATCTGGAATGATGATGATGATG 301  
Qy 281 ACTCATGTACGATCTCTTCCATCTCTGTTGATCAGAAATCAACAACTGGGTCTCCA 340  
Db 302 ACTATGATGAGAAATCTTCCACCTCTCTGATCCGTAATCCGGAATATCTTCAACT 361  
Qy 341 ACTTGCC-----AGGTTACACCATCATGCTGCTCAAGAAACAATCCGGTGGT 391  
Db 362 CCATCTCTGTAACAAATGATACACCATCATCACTGATGGAAGAAATCTTGGTTGA 421  
Qy 392 CCATCGGTATCATCTCAACTTCTGTTGTTTCCCTGATGAGGAGAGAGGACTCCGAGC 451  
Db 422 AAGTATCTGTAACATCGGTGGAATCATCTGAGTCTCTGAGGACACTCAGGAATCAAAAC 481





QY 772 AACGTCGTCAGGACTACTGGGTAAACGACCTGAGATACAAACAGGAGTACTACATGGTC 831  
DB 709 GGTATCCTGAAGAGACTTGTGGGTGACTACTGCTGAGTACGACAAACCGTACTACATGTCG 768  
QY 832 AACATCGACTACTTGAACAGATACATATGACGCAAC 867  
DB 769 AATCTGTACGATCGACAAATAGTTGACGTCAAC 804

## RESULT 5

US-08-405-496A-22  
; Sequence 22, Application US/08405496A  
; Patent No. 5919665  
; GENERAL INFORMATION:  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM  
; TITLE OF INVENTION: NEUROTOXIN  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/405,496A  
; FILING DATE: 16-MAR-1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: INGOLIA, DIANE E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPND-01308  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1330 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: Double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1314  
US-08-405-496A-22

Query Match 12.3%; Score 168.8; DB 2; Length 1330;  
Best Local Similarity 54.8%; Pred. No. 1.1e-28;  
Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;  
QY 64 ATCATCAAGAGTACTCAACAACATCAAGACTCCAGATCCTGTCCTGCGAGAACCGT 123  
DB 19 ACCCTTCACTGATACATCAAGACATCATCAATCTCCATCTGACCTGCGCTACGAA 78

QY 124 AAGAACACCTTGGTGTGACACCTCGGTTTACCAACGCCGAGGTCTCCGAGAGGGTGACGTC 183  
DB 79 TCCAATCACCTGTAGTCAGACCTGTCTCGTACGCTTCCAAAATCAACATCGGTTCTTAAAGTT 138  
QY 184 CAGCTGAACCAATCTTCCCATTCGACTTCAAGCTGGGTTCCCTCCGGTGAGGACAGAGT 243  
DB 139 AACTTCGATCGATCGACAAAGATCAGATCCAGCTG-----TTCAATCTGGATCTTCC 192  
QY 244 AAGTCTATCGTCAACCCAGAGAGAACATCGTCTACAATCTCCATGAGTACGAGTCTTCTCC 303  
DB 193 AAAATCGAAGTTATCTCTGAAGAATGCTATCGTATACAACATCTATGACGAAAATCTCTCC 252  
QY 304 ATCTCCTCTGGATCAGAATCAACAAGTGGTCTCCAACCTTGC-----AGGTTAC 354  
DB 253 ACCCTCTCTGGATCGCTATCCCGAATATCTCAATCTCAATCTCTGTGACAAATGAATAC 312  
QY 355 ACCATATCGACTCCGTCAGAACAACCTCGGTTGGTTCATCGTATCATCTCCAAATTC 414  
DB 313 ACCATATCAACTGCATGGAACAACATCTGTTGGAAAGTATCTCTGAATACGAGTAA 372  
QY 415 CTGCTCTTCAACCTGAGCAGACGAGGACTCCGAGCAGTCCCATCAACTCTCTCTACGAC 474  
DB 373 ATCATCTGGACTCTGAGGACACTCAGGAATCAACAGCGGTGTTGATTCAAATACTCT 432  
QY 475 ATCTCCAACACGCTCTCTGTTTACACAA-----GTGGTTCTTCTGCTACCGTCAACCAAC 531  
DB 433 CAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTTCTGTTACCATCAACCAAT 492  
QY 532 ATGATGGTAACTGAAGATCTACATCAACGGTAAAGCTGATCGACACCATCAAGTCAAG 591  
DB 493 CGTCTGAATAACTCCAAAATCTACATCAAGCGCGCTCTGATCGACAGAAACCGATCTCC 552  
QY 592 GAGTTGACCGGTATCAACTTCTCCAAGACCATCACCTTTCGAGATCAACAAGATCCAGAC 651  
DB 553 AATCTGGTAACTCCAGCTTCTTAATCAATCATGTTCAAACTGGACGGTGTCTGAGAC 612  
QY 652 ACCGGTCTGATCACTCCGACTCCGACCAACATCAAGATGAGTGGATCGTACTCTACATC 711  
DB 613 ACTC-----ACCGTACATCTGGAATCAAAATACTCAATCTG 648  
QY 712 TTGCGCAAGAGTTGGACGGTAAAGGACATCAACATCTGTTCAACTCTTGCAGTACAC 771  
DB 649 TTGCAAAAGAACTGAGCAAGAAAGAAATCAAGACCTGTACGACAAACAGTCCATCT 708  
QY 772 AACGTCTGAGGACTACTGGGTGAGGAGTACGATGAGTACAAAGAGGAGTACATGGTC 831  
DB 709 GGTATCCTGAAAGACTTCTGGGTGACTACTGCTGAGTACGACAAACCGTACTACATGTC 768  
QY 832 AACATCGACTACTTGAACAGATACATGTCAGCAAC 867  
DB 769 AATCTGTACGATCGACAAATAGTTGACGTCAAC 804

## RESULT 6

US-08-915-136-22  
; Sequence 22, Application US/08915136  
; Patent No. 5290960  
; GENERAL INFORMATION:  
; APPLICANT: KINK, JOHN A.  
; APPLICANT: THALLEY, BRUCE S.  
; APPLICANT: PADHYE, NISHA V.  
; APPLICANT: FIRCA, JOSEPH R.  
; APPLICANT: STAFFORD, DOUGLAS C.  
; TITLE OF INVENTION: VACCINE AND ANTI-TOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDLEN & CARROLL, LLP  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA

ZIP: 94104  
COMPUTER TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/915,136  
APPLICATION NUMBER: US/08/915,136  
FILING DATE: 02-DEC-1993  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/480,604  
FILING DATE: 16-MAR-1995  
APPLICATION NUMBER: US 08/405,496  
FILING DATE: 04-DEC-1992  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: INGOLIA, DIANE E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-01763  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1330 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..1314  
US-08-915-136-22

Query Match 12.3%; Score 168.8; DB 4; Length 1330;  
Best Local Similarity 54.8%; Pred. No. 1.1e-28;  
Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;

QY 64 ATCATCAGGAGTACTTCAACAACATCAACGACTCCAGATCCCTGCTCCCTCGAGACCGT 123  
DB 19 ACCTTCACTGAATCATCAAGAATCATCAATACCTCCATCTGAACTCGCGTACGAA 78

QY 124 AAGAACACCTTGGTGCAGACCTCCGGTTACAAACCGCGAGGTCTCCGAGAGGGTGCAGTC 183  
DB 79 TCCAAATCACTGATCGACCTGTCTCGCTACGCTTCCAAATCAACATCGTCTTAAAGTT 138

QY 184 CAGCTGAACCAATCTTCCCATTCGACTTCAAGTGGTTCCTCCGGTGGAGACAGGT 243  
DB 139 AACTTCGATCCGATCGAAGAATCAAGATCGATCGAGTGTGCTTCAATCTGGAATCTTC 192

QY 244 AAGGTTCATCGTCAACCCAGAGAACTCGCTTCAACATCCATGTACGAGTCTCTCC 303  
DB 193 AATATCGAAGTATCTTGAAGATCTATCTATACACTCTATGTACGAAACTTCTCC 252

QY 304 ATCTCTTCTGGATCAGATCAACAAAGTGGTCTCCAACTTGC-----AGGTAC 354  
DB 253 ACCTCTTCTGGATCGGTATCCGAAATCTCAACTCCATCTCTGTGAACAATGATAC 312

QY 355 ACCATCATCGACTCCGTCAAGAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTC 414

Db 313 ACCATCATCACTGCATGGAAACAATCTGGTTGGAAAGTATCTCTGAACCTACGGTGA 372  
QY 415 CTGGTCTTCACTCCCTGAAGCAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGAC 474  
Db 373 ATCATCTGGACTCTGCAGGACACTCAGGAAATCAAAACAGCGGTGTGTATTAATAACTCT 432  
QY 475 ATCTCCAAACACGCTCTCGTGTACACAA--GTGGTTCTTCTGTCACCGTCCACCAAC 531  
Db 433 CAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTCTGTACCATCAACAAT 492  
QY 532 ATGATGGGTAAACATGAAGATCTACATCAACGTTAAGTATGATCGACACCATCAAGGTCAAG 591  
Db 493 CGTCTGAATRACTCCAAATCTATCATCAACGCGCTGTGATCGACCAAGAAACGATCTCC 552  
QY 592 GAGTGGACCGGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAAGATCCCGAC 651  
Db 553 AATCTGGGTAACTCCACGCTTCTAATAACATCATGTCAAACCTGGACGGTGTGCTGAC 612  
QY 652 ACCGGTCTGATCACTCCGACTCCGACATCAACATCAACATGTGGATCGGTGACTTCTACATC 711  
Db 613 ACTC-----ACCGTACATCTGGATCAATACATCTCAATCTG 648  
QY 712 TTCGCCAAGAGTTGGACGTTAAGGACATCAACATCTCTTCAACTCTTGCAGTACACC 771  
Db 649 TTCGACAAAGAACTGAACGAAAGAAATCAAGACCTGTACGACCAACCACTCCATCTT 708  
QY 772 AACGTCTGTAAGGACTACTGGGGTAAAGCCTGAGATACAAACAAGGAGTACTACTATGTC 831  
Db 709 GGTATCCTGAAAGACTTCTGGGTGACTCTCTGAGTACGACAAACCGTACTACTATGTCG 768

QY 832 AACATCGACTACTTGAACATGATCATATGATCGGCAAC 867  
Db 769 AATCTGTACGATCCGAAACAATACGTTGACGTCAAC 804

RESULT 7  
US-08-700-651-1  
; Sequence 1, Application US/08700651B  
; Patent No. 6015682  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN  
; APPLICANT: LEECH, JAMES  
; APPLICANT: NELSON, RICHARD, C.  
; APPLICANT: GUT, JIRI  
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
; TITLE OF INVENTION: INFECTIONS  
; FILE REFERENCE: 480.19-4(HV)  
; CURRENT APPLICATION NUMBER: US/08/700,651B  
; EARLIER FILING DATE: 1997-08-14  
; EARLIER APPLICATION NUMBER: 08/415,751  
; EARLIER FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 5163  
; TYPE: DNA  
; ORGANISM: Cryptosporidium parvum  
US-08-700-651-1

Query Match 9.9%; Score 136; DB 3; Length 5163;  
Best Local Similarity 50.6%; Pred. No. 2.9e-21;  
Matches 328; Conservative 0; Mismatches 320; Indels 0; Gaps 0;

QY 541 AACATGAAGATCTACATCAACGGTAAAGTATGATCGACACCATCAAGGTCAAGGTGAC 600  
DB 577 AACACACACACACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 636  
QY 601 GGTATCACTTCTCCAGACCATCACTTTCGAGATCAACAAGATCCAGACACCGGTCTG 660  
DB 637 AACAAACACTACTACACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 696  
QY 661 ATCACTCCGACTCCGACACATCAACATGTGGATCGGTGACTTCTACTACTTCTCGCCAAG 720







```

; MOLECULE TYPE: DNA (genomic)
US-08-928-361B-1

Query Match
Best Local Similarity 8.4%; Score 115.4; DB 3; Length 7334;
Prior Application Data: 49.0%; Pred. No. 1.2e-16;
Matches 308; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 606 CAATCTCCAGACATACCTTCGAGATCAACAGATCCAGACACCGGTCTGATCAC 665
DB 2621 CTACTACTAGCAACAAACAAACAAACAGCAACAACTACTACTACTACTACTACTA 2680

QY 666 CTCGACTCCGACAAATCAACATGTGGATCGTGACTTCTACATCTTCGCGCAAGGAGTT 725
DB 2681 CTACTACTAGCAACAAACAACTACTACTAGCAACAACTACTACTACTACTACTA 2740

QY 726 GGACGTAAGGACATCAATCCTGTTCAATCTCTTGCAGTACACCAACGTCGTCAGGA 785
DB 2741 CTACACCAACAACTCAACCAACCAACCACTACCAACCAACCACTACCAAGAAC 2800

QY 786 CTACTGGGTAAAGCACTGAGATCAACAAAGGAGTACTACTGTTCAACATCGACTACTT 845
DB 2801 CAACAAACAAACAAACAAACAACTACTACTACTACTACTACTACTACTACTACTA 2860

QY 846 GACAGATACATGTAGCGCACTCAGACAGATCGTCTTCAACACCAACGATACACAA 905
DB 2861 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 2920

QY 906 CGACTTCAACGAGGTTTACAGATCATCAAGCGTATCAGAGGTAACACCAACGACAC 965
DB 2921 CCACACTCCAGAGAACCAACAAACAAACAAACAAACAACTACTACTACTACTA 2980

QY 966 CAGAGTCAGAGGTGGTACATCCTGTACTTTCGATGATGATATCAACAAACAGGCTTCAA 1025
DB 2981 CCAGCAACAAACAAACGAGCAACCAACCAACCACTACTACTACTACTACTACTA 3040

QY 1026 CCGTTCATGAAGACGAGACATGTACGCGACACACCACTCCACCGAGGACATCTAGC 1085
DB 3041 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 3100

QY 1086 CATCGGTCTCGTGAGCAACCAAGGACATCAACGACAAACATCATCTTCCAGATCCAGCC 1145
DB 3101 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTA 3160

QY 1146 AATGAACACACTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAAGGTGAGAA 1205
DB 3161 CAACAAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 3220

QY 1206 CATCTCCGGTATCTGTTCCATCGGTACCT 1234
DB 3221 CCACAAACAACTACTACTTCTGAACT 3249

RESULT 13
US-08-961-527-363
; Sequence 363, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 363:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-363

Query Match
Best Local Similarity 6.9%; Score 95; DB 4; Length 4483;
Matches 558; Conservative 0; Mismatches 665; Indels 24; Gaps 4;

QY 29 TCTTCTCTTACACCAACTCCCTGTTTGAAGGACATCATCAACGAGTACTTCAACAACA 88
DB 2362 TCAGGCTCTGAATCGCATCAACGAGTGGCTCAGCCTCAGCAAGTACTAGTGCATCAGCA 2421

QY 89 TCAACGACTCCAGATCTCTGTCCTTGCAGAACCGTGAAGAACACCTTGGTCGACACTCG 148
DB 2422 TCAGCATCAACGAGTGCATCGGCTTTCAGCAAGTACCAAGGCTCAGCTTCAGCAAGCACC 2481

QY 149 GTTACAACCCCGAGGTCTCCGAGAGGAGGTGAGCTCCAGCTGAACCCCAATCTTCCCATTCG 208
DB 2482 AGTGGCTCAGCTCAGCAAGTACCAAGGCTCAGCTTCAGCAAGCACCAGTGCCTCAGCT 2541

QY 209 ACTTCAAGCTGGGTTCTCCGGTGAGGACAGAGGTAGGTATGTCACCCAGAACAGAGA 268
DB 2542 TCAGCAAGTACCAAGTGCCTGTCAGCTTCAGCTTCAGCAAGTGGCTTCAGCAAGTACC 2601

QY 269 ACATCTCTTACAACTCCATGTACGAGTCTCTCCATCTCTTCCATCTCTTGGATCAGAAATCAACA 328
DB 2602 TCAGCTCTGATCAGCATCAACGAGTGCATCAGCTTCAGCATCAACAGTGTCTCAGCT 2661

QY 329 AGTGGGTCTCCAACTTCCAGGTTACACCATCATCGATCGCTCCAGAACAACTCCGGTT 388
DB 2662 TCAGCAAGTACCAAGTGCCTGTCAGCTTCAGCATCAACGAGTGTCTCAGCTTCAGCTCAACC 2721

QY 389 GGTCCATCGGTAT---CATCTCCAACTTCTCTGGTCTTCACTCCCTGAAGCAGAACGAGGACT 445
DB 2722 AGTGGCTCTGATCAGCATCAACAGTGCCTCGGCTTCAGCAAGCACCAGTGCCTCGGCT 2781

QY 446 CCGAGCAGTCCATCAACTTCTCTAGCAGATCTCCAAACAGCTCCCTGGTTTACAAACAGT 505
DB 2782 TCAGCAAGTACTAGTGCATCGGCTTCAGCATCGACAGTGTCTGATTCGGCATCAACG 2841

QY 506 GGTCTCTCTGCTACCGTTCACCAACA---CATGATGGTAAACATGAAGATCTACA 556
DB 2842 AGTGGCTTCGCTTACGATCAACAGTGCCTCGGCTTCAGCAAGCACCAGTGCCTCGGCT 2901

QY 557 TCAACGGTAAAGTGTATCGACACCATCAAGGTCAAGGATTTGACCGGTATCAACTTCTCCA 616
DB 2902 TCTGCATCAACCAAGTGCCTGTCAGCTTCAGCGTCAACAGTGTCTGATTCGGCATCAACG 2961

QY 617 AGACCATCACTTCGAGATCAACAGATCCGAGACACCGGTCTGTGATCAGCTCCGACTCCG 676
DB 2962 AGTGGCTTCGCTTACGATCAACAGTGCCTCGGCTTCAGCAAGCACCAGTGCCTCGGCT 3015

QY 677 ACAACATCAACATGTGATCGGTGATCTTCTACATCTTCGCAAGGAGTGTGACCGGTAAAG 736
DB 3016 TCAGCGCTCAGCTTCGCGCTCAACCAAGTGCCTCGGCTTCAGCAAGCACCAGTGCCTCGGCT 3075

```

Query Match 6.8%; Score 93.4; DB 1; Length 2004;  
Best Local Similarity 44.3%; Pred. NO. 6.1e-12;  
Matches 529; Conservative 0; Mismatches 651; Indels 15; Gaps

Query Match	6.8%	Score 93.4	DB 1	Length 2004	
Best Local Similarity	44.3%	Prod. No. 6.1e-12			
Matches 529	Conservative	0	Mismatches 651	Indels 15	Gaps
QY	51	CCTGTTGAAGGACATCATCAACGAGTACTTTCACACATCATCAACGACTCCAAAGATCCTTC	110		
DB	378	CGTGAGGCGCGCATCGGTCCCAAGGCGATCAGCTTCGSCGTGAGCGTGAACACTACCAGCA	437		
QY	111	CTGCGAGAACCGTAAGACACACCTTGTGCGACACCTCCGGTTACACGCCGAGGTCCTCCGA	170		
DB	438	CASCGAGACCGTGGCCCGCAGAGTGGGGCACCGACCGCGAACACCGACCGCAAGTTTCAACAC	497		
QY	171	GGAGGGTGACGTCCAGCTGAACCCAAATCTTCCCATTCGACTTCAAG---CTGGGTTCCCTC	237		
DB	498	CGCCAGCCCGGGCTACCTGAACGCCCAACGTGGGCTACACACACGTGGGCGCACCGGCGCAT	557		
QY	228	CGGTGAGGACAGAGTGAAGTTCATCTGTCACCCAGAACGAGAACATCGTCTACAACTCCAT	287		
DB	558	CTACGACGTGAAGCCCGCACCCAGCCTCGTGTGTAACACGACACCATCGCCACCATCAC	617		
QY	288	GTACGAGTCTCTTCATCTCTCTCTGGATCAGATCATCAACAGTGGGTCTCTCAACTTGGC	347		
DB	618	CGCCAACTGCGAATTCACACGCCCTGAACTACGSCCGCGGAGAGCTACCCCAAGAAGGG	677		
QY	348	AGGTTTACACCATCATCGACTCCGTCGAAGAACAACTCCGGTGGTCCATCGGTATCATCTC	407		
DB	678	CCAGAACGGCATCGCCATCACCGATGGACGACTTCAACAGCCACCCCATCCCTGAA	737		
QY	408	CBACTTCTGTGCTTCAACCTGGAAGCAGACAGGAGACTCCGAGCAGTCCCATCACTTCTC	467		
DB	738	CAAGAAGCAGGTGGACAACTGTGTAAACAAAGCCCATGATGCTGGAGACCAACAGAC	797		
QY	468	CTACGACATCTCCAAACAGCTCCTGTTTACAACAGTGGTTCCTGCTGACCGGTCAACAA	527		
DB	798	CGACGCGCTCTTACAAGATCAAGGACACCCACCGCGCAACATCTGTACCGCGCGGAGTGGAA	857		
QY	528	CAACATGATGGTGAACATGAAGATCTACATCAACGGTAAGCTGATCTGACACCATCAA--	584		
DB	858	CGGCGTGTACAGCAGATCAAGCCCAAGACCCCGCCAGCATCATGCTCGACGACGGCAGCG	917		
QY	585	---GGTCAAGGAGTTGACCGGTATCAACTTCTTCCAAGACCATCATCTTCGAGATCAACAA	641		
DB	918	CGTGGCCGAGAAGCGCGTGGCCCGCAAGGACTACGAGAACCOCGAGGACAAAGCCCCCAG	977		
QY	642	GATCCGACACACCGGCTGTATCACTTCCGACTCCGACCAACATCAACATGTGGATTCGGTGA	701		

Db 978 CCTGACCTGAGAGGAGGCGCTGAGCTGAGCTACCCGACGAGATCAAGGAGATCGAGG 1037  
QY 702 CTTCTATATCTTCGCAAGGAGTTGGAGGTAAGGAGATCAACATCTCTGTTCAACTCCTT 761  
Db 1038 CCGTGTGTACTACAAGAACAGCCCATCTACGAGAGCAGCGTGATGACCTATCTAGACGA 1097  
QY 762 GCAGTACACCAACGTCGTCAGGAGTACTGCGGTAAAGCACTGAGATACACAGGAGTA 821  
Db 1098 GAACACCGCAAGGAGGTGACCAAGCACTGAAAGACACCAAGGAGTTCAGGAGCGT 1157  
QY 822 CTACATGTCACATCGACTACTTGAACAGATATACATGTAACCCCACTCCAGACAGATCGT 881  
Db 1158 GAGCCACCTGTAGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1217  
QY 882 CTTCAACACAGAGTACACCAAGCACTTCAAGAGGTTTACAGAGTATCATCAAGCG 941  
Db 1218 CCGTGTACGACACGCGGAGAGCAAGCAAGCACTCGGCAAGTGGACCAACCAACAT 1277  
QY 942 TATCAGAGGTAAACCAACGACACAGAGTCAAGAGTGGTGACATCTCTGTTCTGAGAT 1001  
Db 1278 CGTGAGCGGCGACACAGCGGCAAGAGCAGTACAGAGCAACACCCGAGCGCACT 1337  
QY 1002 GACTATCAACAGAGGCTACACCTGTTATGATGAAGCAAGAGCACTGTACGCGCAAA 1061  
Db 1338 GACCCCTGAACCGGAGCGGCGGAGGAGCT-----GAACAAGAACCGGAGTACTAT 1391  
QY 1062 CCACTCCAGGAGGAGTACCTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121  
Db 1392 CAGCCTGTACATGAGAGGCGGAGAGCAACCGAGTCCGAGTACCATCGAGCGGAGAT 1451  
QY 1122 CAACATCATCTTCCAGATCCAGCCATGAACCAACATTTACTACTACGTTCCCATCTT 1181  
Db 1452 ATACCCCATCACCACCAAGACCGTGAACGCTGAACCAAGCAACATACAAGCGCTTGGAT 1511  
QY 1182 CAAGTCCACTCAAGGAGTGAAGCACTCCGCTATCTGTTCCATCGTACCTAC 1236  
Db 1512 CATGCCCAACATCAAGAGCAACCCCATCAGAGCGCTGCACATCAAGACCAAC 1566

## RESULT 15

US-08-471-044-18  
; Sequence 18, Application US/08471044  
; Patent No. 5840868  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Gregory W  
; APPLICANT: Koziel, Michael G  
; APPLICANT: Mullins, Martha A  
; APPLICANT: Nye, Gordon J  
; APPLICANT: Carr, Brian  
; APPLICANT: Desai, Nalini M  
; APPLICANT: Kostichka, N. Kristy  
; APPLICANT: Duck, Nicholas B  
; APPLICANT: Estruch, Juan J  
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30B  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,044  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/463,483  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/314,594  
; FILING DATE: 09-SEP-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,018  
; FILING DATE: 23-MAR-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/037,057  
; FILING DATE: 25-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pace, Gary M.  
; REGISTRATION NUMBER: 40,403  
; REFERENCE/DOCKET NUMBER: CEC 1695/CIP3/DIV6 - SOLV3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8582  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2004 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; NAME/KEY: misc\_feature  
; LOCATION: 1..2004  
; OTHER INFORMATION: /note= "Maize optimized DNA  
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"  
US-08-471-044-18  
  
Query Match 6.8%; Score 93.4; DB 2; Length 2004;  
Best Local Similarity 44.3%; Pred. No. 6.1e-12;  
Matches 529; Conservative 0; Mismatches 651; Indels 15; Gaps 3;  
  
QY 51 CCGTGTGAAGGACATCATCAACGAGTACTTCAACAACATCAACGACTCCAGAGTCTCTGTC 110  
Db 378 CCGTGAAGCGCGCATCGGTCCCAAGGCGATCAGCTTCGGCGTGAGCTGAATACCAAGCA 437  
QY 111 CCGTGAGACACCGTAAGAACACCTTGGTGCACACCTCCGGTTTACACCGCGAGGTCTCCGA 170  
Db 438 CAGGAGACCGTGGCGCCAGGAGTGGGGCACCGACCGGCAACACACGAGTCAACAC 437  
QY 171 GGAGGTGACGTCCAGCTGAACCCAAATCTTCCATTGCACTTCAAG---CTGGTTCCTC 227  
Db 498 CGCCAGCGCGCGCTACCTGAACGCCAACGTCGCTACAACAACGTTGGCCACCGCGCAT 557  
QY 228 CCGTGAGGACAGAGTAAGTCACTGTCACCCAGACAGAGACATGCTCTACACTCCAT 287  
Db 558 CTACGACGTGAAGCCCGCCACCGATCGTTCGTGAACAACGACACCATCGCCACCATCAC 617  
QY 288 GTACGAGTCTCTCTCCATCTCTTCTGGATCAGATCAACAAGTGGGTCTCCCACTTGC 347  
Db 618 CGCCAGTCCGATTCACCGCCCTGAACATCAGCCCGGAGAGCTACCCCAAGAGGG 677  
QY 348 AGGTACACCATCATCATGCTCGTTCAGAAACAATCCCGTTGGTTCATCGGTATCATCTC 407  
Db 678 CCAGAACGCGCATCGCCCATCAGCAGCATGGAGACTTCAACAGCCACCCCATCCCTGAA 737  
QY 408 CAACCTCTCTGGTCTTCAACCTGAAGCAGAGGAGTCCGAGCAGTCCCACTTCTC 467  
Db 738 CAAGAGCAGGTGGAGACACCTGCTGAACAACAAGCCCATGATGCTGGAGACCAACAGC 797  
QY 468 CTACGACATCTCCACAACAGCTTCTGTTTACAACAAGTGGTTCCTGTCACCGTACCAA 527  
Db 798 CGACGCGGTCTACAAGATCAAGGACACCCCGGCAACATCGTGACCGCGCGGAGTGGAA 857  
QY 528 CAACATGATGGGTATCATGACATCTTACATCAACGGTAAGCTGATCGACACCATCA-- 584  
Db 858 CCGCGTGTATCCAGCAGATCAAGGCCCAAGACCCGCGCAGCATCATCTCGTCGACGCGGAGCG 917



Qy	585	---	GGTCAAGGAGTTGACCGGTTATCAACTTCTCCAGAGCCATCACTTCGAGATCAACA	641
Db	918	CGTGGCCGGAAGCGGTGGCCGCAAGGACTACGAGAAGCCCGAGGACAAGACCCCCAG	977	
Qy	642	GATCCCGACACCGGTCTGATCACCTCGGATCCGGAACATCAACATGTTGGATCGGTGA	701	
Db	978	CTGTACCTGAAGGAGCGCCTGAAGCTGAGCTACCCGACGAGATCAAGGAGATCGAGG	1037	
Qy	702	CTTCTACATCTTGGCCAAGGAGTTGGAGGTTAAGGACATCAACATCTGTTCACATCCTT	761	
Db	1038	CTGTCTGTACTACAGAACAAGCCCATCTACGAGAGAGCGTGATGACCTATCTTAGACGA	1097	
Qy	762	GCAGTACACCAACGTTGTCGAAGGACTACTGGGGTAAGCACCTGAGATACAAACAGGATG	821	
Db	1098	GAACACGGCAAGGAGTGAACCAACAGCTGAAGACACCAACCGGCAAGTTCAGGACGT	1157	
Qy	822	CTACATGGTCAACATCGACTTCTGAACAGATACATGTACGCCAACTCCAGACAGATCGT	881	
Db	1158	GAGCCACTGTAGGAGTGAAGCTGACCCCAAGNTGAACGTGACCATCAAGCTGAGCAT	1217	
Qy	882	CTTCAACACGAGACGTAAACAACAAGACTTCAAGAGGGTTTCAAGATCATCATCAAGOG	941	
Db	1218	CCTGTACGACAACGGCGAGAGCAACAGCATCGCGAAGTGGACCAACACCAACAT	1277	
Qy	942	TATCAGAGGTTAACACCAAGCACACAGAGTCAGAGGTGGTGACATCTGTTACATTCCACAT	1001	
Db	1278	CGTCAGCGCGCAACAACGGCAAGACGAGTACAGCAGCAACACCCGACGCCAACCT	1337	
Qy	1002	GACTATCAACAACAGCGCTACAACTGTTGATGAGAACGAGACCATGTACGCGCGNAA	1061	
Db	1338	GACCTGGAACACCGGCGCCAGGAGAGCT-----GAACAAGAACCGCGACTACTACAT	1391	
Qy	1062	CCACTCCACGAGGACATCTAGCCGATCGGTGTCGTTGAGCAGACCAAGGACATCAACGA	1121	
Db	1392	CAGCGTGTACATGAAGAGCGGAAGAACAACCCAGTGCAGATCAACATCGACGCGGAGAT	1451	
Qy	1122	CACATCATCTCCGATCCGCCATGAACACACTTACTACTACGCTTCCCGAGATCTT	1181	
Db	1452	ATACCCCATCCACCACAGACCGTGAACGTGAACAGGACAACATCAAGCGCTTGGACAT	1511	
Qy	1182	CAAGTCCAACTTCAACGGTGGAACATCTCGGTTATCTGTTTCCATCGGTACCTAC	1236	
Db	1512	CATCGGCCCAACATCAAGAGCAACCCCATCAGAGCCTGCACATCAAGACCAAC	1566	

Search completed: November 7, 2002, 13:46:45  
Job time : 74 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 13:24:50 ; Search time 57 Seconds

(without alignments)  
8533.524 Million cell updates/sec

Title: US-09-910-186a-9

Perfect score: 1371

Sequence: 1 gaattcagatgaccatcccc.....tctccgagtaataagaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	1366.2	99.6	1371	10	US-09-350-756-3
2	662	48.3	3876	12	US-10-051-952-4
3	217.4	15.9	1347	10	US-09-350-756-6
4	158.2	14.5	1317	10	US-09-350-756-5
5	168.8	12.3	1338	10	US-09-350-756-1
6	162.4	11.8	1400	10	US-09-350-756-4
7	151.6	11.1	1313	8	US-08-981-087A-6
8	136.6	9.2	702	10	US-09-350-756-10
9	120.4	8.6	1341	10	US-09-350-756-2
10	93.4	6.8	1635	10	US-09-864-761-20241
11	93.4	6.8	1973	10	US-09-864-761-3471
12	92.2	6.7	1397	10	US-09-216-393-343
13	92.2	6.7	1397	10	US-09-216-393-345
14	90.2	6.6	867	10	US-09-216-393-340
15	90.2	6.6	867	10	US-09-216-393-342
16	88.8	6.5	3831	12	US-10-051-952-5
17	82.6	6.0	1959	10	US-09-864-761-4012
18	82.4	6.0	2535	10	US-09-476-242-13
19	81.6	6.0	1075	10	US-09-864-761-19241

c	20	81.6	6.0	1403	10	US-09-864-761-2513	Sequence 2513, Ap
	21	77.6	5.7	2529	10	US-09-476-242-14	Sequence 14, Appl
	22	74	5.4	2298	10	US-09-476-242-22	Sequence 22, Appl
	23	74	5.4	2298	10	US-09-476-242-23	Sequence 23, Appl
	24	74	5.4	2298	10	US-09-476-242-24	Sequence 24, Appl
c	25	73.4	5.4	583	10	US-09-864-761-20772	Sequence 20772, A
	26	71.2	5.2	2523	10	US-09-476-242-15	Sequence 15, Appl
	27	71	5.2	1818	10	US-09-801-368-427	Sequence 427, App
	28	70.8	5.2	2310	10	US-09-476-242-21	Sequence 21, Appl
	29	69.8	5.1	1977	12	US-10-090-624-11	Sequence 11, Appl
	30	69.8	5.1	2541	10	US-09-476-242-12	Sequence 12, Appl
	31	69.4	5.1	2010	12	US-10-032-717-9	Sequence 9, Appl
	32	68.2	5.0	2541	10	US-09-476-242-11	Sequence 11, Appl
	33	67.8	4.9	1518	9	US-09-934-060A-23	Sequence 23, Appl
	34	67.8	4.9	1518	9	US-09-934-060A-29	Sequence 29, Appl
	35	67.8	4.9	1668	9	US-09-934-060A-5	Sequence 5, Appl
	36	67.8	4.9	2159	9	US-09-934-060A-1	Sequence 1, Appl
	37	67.8	4.9	2159	9	US-09-934-060A-3	Sequence 3, Appl
	38	67.6	4.9	1944	10	US-09-864-761-2825	Sequence 2825, Ap
	39	67.4	4.9	2316	10	US-09-476-242-4	Sequence 4, Appl
	40	67.4	4.9	2517	10	US-09-476-242-16	Sequence 16, Appl
	41	67.4	4.9	2517	10	US-09-476-242-17	Sequence 17, Appl
	42	67.4	4.9	2541	10	US-09-476-242-9	Sequence 9, Appl
	43	67.4	4.9	2541	10	US-09-476-242-10	Sequence 10, Appl
	44	66.2	4.8	766	10	US-09-864-761-19608	Sequence 19608, A
	45	66	4.8	2316	10	US-09-476-242-8	Sequence 8, Appl

#### ALIGNMENTS

RESULT 1  
US-09-350-756-3  
; Sequence 3, Application US/09350756  
; Patent No. US20020034521A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP RIID 98-21  
; CURRENT APPLICATION NUMBER: US/09/350,756  
; CURRENT FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092,416  
; EARLIER FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 3  
; LENGTH: 1371  
; TYPE: DNA  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"  
US-09-350-756-3

Query Match	99.6%	Score	1366.2	DB	10	Length	1371
Best Local Similarity	99.8%	Pred. No.	0				
Matches	1366	Conservative	0	Mismatches	3	Indels	0
Gaps	0						
Qy	1	GAATTCAGTACGATCCCATTCACATCTCTCTACACCAACATCCCTGTGTGAAG	60				
Db	1	GAATTCAGTACGATCCCATTCACATCTCTCTACACCAACATCCCTGTGTGAAG	60				
Qy	61	GACATCATCAACGAGTACTTCAACATCAACGACTCCCAAGATCTGTCCTGTGAGAAC	120				
Db	61	GACATCATCAACGAGTACTTCAACATCAACGACTCCCAAGATCTGTCCTGTGAGAAC	120				
Qy	121	CGTAGAACCTTGTGTCGACACTCCGGTTACAAAGCCGAGGTCTCCGAGGGGTGAC	180				

Db 121 CGTAAGAACACCTTGTGTGACACCTCCCGGTTTACACGGCGAGGTCTCCGAGAGGGTGAAC 180  
 QY 181 GTCCAGGTGAACCAATCTTCCCACTTCGACTTCAAGCTGGGTTCCTCCGGTGGAGACAGA 240  
 Db 181 GTCCAGGTGAACCAATCTTCCCACTTCGACTTCAAGCTGGGTTCCTCCGGTGGAGACAGA 240  
 QY 241 GGTAAAGGTGATCTGTCACCCAGACGAGACATCGTCTACAACTCCATGTACAGTCCCTTC 300  
 Db 241 GGTAAAGGTGATCTGTCACCCAGACGAGACATCGTCTACAACTCCATGTACAGTCCCTTC 300  
 QY 301 TCCATCTCTCTTCTGGATGAGATCAACAAAGTGGGTCTCCAACTTCCAGGTTCACCAATC 360  
 Db 301 TCCATCTCTCTTCTGGATGAGATCAACAAAGTGGGTCTCCAACTTCCAGGTTCACCAATC 360  
 QY 361 ATCGACTCCGTCAAGAACAACTCCGGTTCGTCATCCGGTATCATCTCAACTTCCTTGCTC 420  
 Db 361 ATCGACTCCGTCAAGAACAACTCCGGTTCGTCATCCGGTATCATCTCAACTTCCTTGCTC 420  
 QY 421 TTCACCTGAAGCAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCC 480  
 Db 421 TTCACCTGAAGCAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTACGACATCC 480  
 QY 481 AACAAAGCTCTGGTTCACAAAGAGTGGTTCCTGTCACCGTCCACCAACAAATGAGGT 540  
 Db 481 AACAAAGCTCTGGTTCACAAAGAGTGGTTCCTGTCACCGTCCACCAACAAATGAGGT 540  
 QY 541 AACATGAAGATCTACATCAACGGTGAAGTGTGACACCATCAAGTCAAGGATGACC 600  
 Db 541 AACATGAAGATCTACATCAACGGTGAAGTGTGACACCATCAAGTCAAGGATGACC 600  
 QY 601 GGTATCAACTTCTCCAGACCAATCACTTCGAGATCAACAAGATCCCAAGACCGGTCTG 660  
 Db 601 GGTATCAACTTCTCCAGACCAATCACTTCGAGATCAACAAGATCCCAAGACCGGTCTG 660  
 QY 661 ATCACTCCGACTCCGACACATCAACATGTGGTCCGCTGACTTCTACATCTTCCGCAAG 720  
 Db 661 ATCACTCCGACTCCGACACATCAACATGTGGTCCGCTGACTTCTACATCTTCCGCAAG 720  
 QY 721 GAGTTGACCGGTGAAGACATCAACATCTTGTTCAACTCTTCCAGTCAACCAAGTCTGTC 780  
 Db 721 GAGTTGACCGGTGAAGACATCAACATCTTGTTCAACTCTTCCAGTCAACCAAGTCTGTC 780  
 QY 781 AAGGACTACTTGGGTGAAGACATCAACATCTTGTTCAACTCTTCCAGTCAACCAAGTCTGTC 840  
 Db 781 AAGGACTACTTGGGTGAAGACATCAACATCTTGTTCAACTCTTCCAGTCAACCAAGTCTGTC 840  
 QY 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACAGACGTAAC 900  
 Db 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTCAACACAGACGTAAC 900  
 QY 901 AACACGACTTCAACGAGGTTCACAGATCATCATCAAGCGTATCAGAGGTAAACCAAC 960  
 Db 901 AACACGACTTCAACGAGGTTCACAGATCATCATCAAGCGTATCAGAGGTAAACCAAC 960  
 QY 961 GACACAGAGTCAAGAGTGGTGAATCTTCTGACATCTTCTGACATGACTATCAACAAAGGCC 1020  
 Db 961 GACACAGAGTCAAGAGTGGTGAATCTTCTGACATCTTCTGACATGACTATCAACAAAGGCC 1020  
 QY 1021 TACAACTGTTTCATGAAGAACGAGACATGTAGCGGACACCACTCCACGAGGACATC 1080  
 Db 1021 TACAACTGTTTCATGAAGAACGAGACATGTAGCGGACACCACTCCACGAGGACATC 1080  
 QY 1081 TACGCCATCGGTCTGGTGAGACCAAGGACATCAACGACAAATCATCTTCCAGATC 1140  
 Db 1081 TACGCCATCGGTCTGGTGAGACCAAGGACATCAACGACAAATCATCTTCCAGATC 1140  
 QY 1141 CAGCCATGAACAACTTACTACTACTAGGTTCGCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
 Db 1141 CAGCCATGAACAACTTACTACTACTAGGTTCGCCAGATCTTCAAGTCCAACTTCAACGGT 1200  
 QY 1201 GAGAAATCTCCGGTATCTGTTCCATCTCGTACCTACAGATTCCTGCTGGTGGTCACTGG 1260  
 Db 1201 GAGAAATCTCCGGTATCTGTTCCATCTCGTACCTACAGATTCCTGCTGGTGGTCACTGG 1260

QY 1261 TACAGACACAACACTACTTGTTCCTCAACTGTCAAGCAGGTAACCTCCTCTCTGCTGGAG 1320  
 Db 1261 TACAGACACAACACTACTTGTTCCTCAACTGTCAAGCAGGTAACCTCCTCTCTGCTGGAG 1320  
 QY 1321 TCCACTTCCACCCACTGGGGATTCGTCCTCCAGTCTCCGAGTATAGGAATTC 1371  
 Db 1321 TCCACTTCCACCCACTGGGGATTCGTCCTCCAGTCTCCGAGTATAGGAATTC 1371

RESULT 2

US-10-051-952-4  
 ; Sequence 4, Application US/10051952  
 ; Patent No. US20020107199A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Walker, Patricia  
 ; TITLE OF INVENTION: Methods of Administering Botulinum Toxin  
 ; FILE REFERENCE: 2933CIP  
 ; CURRENT APPLICATION NUMBER: US/10/051,952  
 ; CURRENT FILING DATE: 2002-01-17  
 ; PRIOR APPLICATION NUMBER: 09/730,237  
 ; PRIOR FILING DATE: 2000-12-05  
 ; NUMBER OF SEQ ID NOS: 12  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4  
 ; LENGTH: 3876  
 ; TYPE: DNA  
 ; ORGANISM: botulinum toxin  
 US-10-051-952-4

Query Match 48.3%; Score 662; DB 12; Length 3876;  
 Best Local Similarity 68.1%; Pred. No. 5.3e-156;  
 Matches 920; Conservative 0; Mismatches 430; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACCATCTCTCTACACCAACAACTCCCTGTTCAAGGACATCATCAAC 72  
 Db 2527 ACAATACCCTTTAATATTTTTCATATATCTAATAATCTTTATTAAGACATATATTAAT 2586  
 QY 73 GAGTACTTCAACAACATCAACGACTCCAAGATCTCTCCCTGCAGAACCGTAAGAACC 132  
 Db 2587 GAATATTTCAATAATATTAAATGATTCAAAAATTTTGAGCTCTACAAAACAGAAAAAATACT 2646  
 QY 133 TTGGTCGACACCTCCGGTTACAAACCCGAGGCTCCGAGGAGGTGAGTCCAGCTGAAC 192  
 Db 2647 TTAGTGGATACATCAGGATATATCGAAGTGAAGTGAAGGAGGTGAGTCCAGCTTAAT 2706  
 QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTCTCCCTCCGGTGAAGACAGAGTAAAGTCAATC 252  
 Db 2707 CCAATATTTCCATTTGACTTTAAATTAGGTAGTTCCAGGGAGGATAGAGGTAAGATTATA 2766  
 QY 253 GTACCCCAAGACGAGACATCTCTACACCTCCATGTACGAGTCTCTCTCCATCTCCTTC 312  
 Db 2767 GTAACCCGAATGAATAATTTGTATATATCTATATATCTATGATAAGGTTTAGCATAGTTT 2826  
 QY 313 TGATCAGATCAACAAGTGGGTCTCCAACTTGCAGGTTACACCATCATCGACTCCGCTC 372  
 Db 2827 TCGATTAGATTAATAAATGGGTGAAGTAAATTTACCTGGATATCTATATAATGATAGTTT 2886  
 QY 373 AAGAACAACTCCGGTGGTTCATCGGTATCATCTCCAACTCTCCCTGGTCTTCCACCTGAAG 432  
 Db 2887 AAAAATAAATAAATGTTTTTTTGTAACTGTTACTAACTATATAGTATATTTTAGTATTTTACTTTAAA 2946  
 QY 433 CAGAACGAGGACTCCGAGCAGTCCCATCAACTCTCTCTACGACATCTCCAAACAGCTCCT 492  
 Db 2947 CAATATGAGATAGTAGAACAAGTATAAATTTAGTTATGATATATCAAAATAGTCTCCT 3006  
 QY 493 GGTACAACAAGTGGTTCCTTCCTCCCGTCAACCAACACATGATGGGTAAACATGAAGATC 552  
 Db 3007 GGATAGATAAATAAATGTTTTTTTGTAACTGTTACTAACTATATGATGGGAAATATGAAGATT 3066  
 QY 553 TACATCAACGTAAGTGTATCCGACCATCAAGGTCAGGAGTTGACCGGTATCAACTTC 612  
 Db 3067 TATATAATGGAATAATTAATAGTATCTATAAAGTTAAGAACTAACTGGAATTAATTTT 3126

```
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 6
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-6

QY 613 TCCAGAGCCATCCTTCGAGATCAACAAGATCCCGACACACCGGTCTGTATCCCTCCGAC 672
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3127 AGCAAACTATAACATTGAAATAAAATAAAATCCAGATACCGGTTGATTACTTCAGAT 3186

QY 673 TCCGACACATCAACATGGATCGGCTGACTTCTACATCTCCGCAAGGAGTTGGACGGT 732
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3187 TCTGATACATCAATATGTGGTAAGAGATTTTATATATTTGCTAAAGAAATTAGATGT 3246

QY 733 AAGGACATCAACATCCCTGTTCAACTCTTCGAGTACACCAACGCTGCTCAAGGACTACTGG 792
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3247 AAGATATTAATATATTATTTAATAGCTTGCATATACATAATGTTCTAAAGATTATTGG 3306

QY 793 GGTAAACACCTGAGATACACCAAGGAGTACTATAGTGTCAACATCCACTACTTGAACAGA 852
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3307 GGAATGATTTAAGATATAAAGAATAATTATGTTTAATATAGATTATTAAATAGA 3366

QY 853 TACATGTACGCCAACTCCAGACAGATCGTCTCAACACACAGAGCTGAACAAACAGACTTC 912
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3367 TATATGTATGCGAACTCAGCAAAATTTGTTTTTAATACAGTAGAATAATAATGACTTC 3426

QY 913 AACGAGGTTACAGATCATCATCAACGCTATCAGAGGTTAACCAACACAGCAGAGTC 972
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3427 AATGAAGGATATAAATATAATAAAGAATCAGAGGAATAACAAATGATAGAGTA 3486

QY 973 AGAGTGTGTGACATCCGTACTTCGACATGACTATCAACAACAAGGCTTACAACTGTTTC 1032
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3487 CGAGGAGGATATTTATATTTTGTATGACAAATTAATAACAAGCATATAATTTGTTT 3546

QY 1033 ATGAAGAAGAGACCATGTACGCGGCAACCACTCCACCGAGGACATCTACGCCATCGGT 1092
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3547 ATGAAGAATGAACATATGTATGCAATTAATCATAGTACTGAAGATATATATGTTAGT 3606

QY 1093 CTGGTGTAGCAGACCAAGGACATCAACGCAACATCATCTTCCAGATCCAGCCAAATGAAC 1152
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3607 TTAAGAGAACAAACAAAGGATATAATGATTAATATATATTTCAATATCAACCAATGAAT 3666

QY 1153 AACACTTACTACTACGTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCC 1212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3667 AATACTTATTTATAGGATCTCAAAATTTTAAATCAAAATTTTAAATGAGAAATATTTCT 3726

QY 1213 GGTATCTGTCCTACGTTACCTACAGATTCGCTGCTGGTGTGCTGCTGCTGCTGCTGCTGCT 1272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3727 GGAATGTTTCAATAGTACTTATCTGTTAGACTGGAGTGATTTGGTATAGACAAAT 3786

QY 1273 TACTTGTTCACATGTCAAGCAGGTTACTACGCTCTCTGTTGTGGAGTCCACTTCCACC 1332
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3787 TATTTGGTCCCTACTGTGAAGCAAGGAAATATGCTTCAATATTATTAATCAATCAACT 3846

QY 1333 CACTGGGATTCGTCGCGGATCTCCGAGTAA 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3847 CATTTGGGTTTGTACCTGTAAGTGAATAA 3876
```

## RESULT 3

```
US-09-350-756-6
; Sequence 6, Application US/09350755
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
```

```
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 6
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-6
```

```
Query Match 15.9%; Score 217.4; DB 10; Length 1347;
Best Local Similarity 52.5%; Pred. No. 4e-45;
Matches 631; Conservative 0; Mismatches 516; Indels 54; Gaps 5;

QY 43 AACAACTCCCTGTTGAGGACATCATCAAGAGTACTTCAACAACATCAACGACTCCCAAG 102
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 AAGACACCACTCCCTGATGCCAGGCTTCAACAACATCATCTCCAACTATCTCCCAAGGCC 63

QY 103 ATCTGTCCCTGCGAGACCCGTAAGAACACCTTGTGTGCACACCTCCCGTTTACACGCCGAG 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 ATCTGTCCCTGCTCTACCGTGGTGTGCTGATCGACTCTCCCGTTTACGAGGCCACC 123

QY 163 GTCTCCGAGGAGGTGACGTCCAGCTGAACCCAAATCTTCCCATTCGACTTCAGCTGGGT 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 ATCAACCTCGGTTCCGACGTCATCTTCAACGACATCGGTAAACGTCAGTTCAAGCTGAAC 183

QY 223 TCCCTCCGTTGAGGACAGAGGTAAAGGTTCATCGTCACCCAGAGAGAGAACATCGTCTACAAAC 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 AACTCC-----GAGAACTCCAACATCAGCGCCACCAGTCCCAAGTTCGTCGTACGAC 237

QY 283 TCCATGTACGAGTCCCTTCTGATCTCTGATCAGATCAGATCAACAAGTGGGTCTCCCAAC 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 TCCATGTTCGACAACTTCTCCCACTAACTTCTGGTCCGTAACCCCAAGTACAAACAAAC 297

QY 343 -----TTGCCAGGTTACACCATCATCGACTCCGTCGTCGTCGTCGTCGTCGTCGTCGTC 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 GACATCCAGACCTACCTGCGAGACGAGTACACCATCATCTCTCTGTATCAAGACGACTCC 357

QY 385 GGTGTGTCATCGGTATCATCTCCAACTTCTGTTCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 GGTGTGAGAGTCTCCATCAAGGAAACCGGTATCATCTGGACCCGTGATCGACTCCCAAGGCC 417

QY 445 TCCGAGGACGTCATCAACTTCTCTACGACATCTCCAAACAACTCTCTCTCTCTCTCTCTCTCT 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 AAGTCCAAAGTCCATCTCTTGGAGTACTCCATCAAGGACAAACATCTCCGACTACATCAAC 477

QY 502 AAGTGTGTTCTGTCACCGTCACCAACAAATGATGGGTAAATGATGATGATGATGATGATGATGAT 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 AAGTGTGTTCTCCATCAACCATCAACCAAGCCGCTGGTAAACGCCAACATCTACATCAAC 537

QY 562 GGTAAAGTGTGTCGACACCATCAAGGTCAAGGAGTTGACGGGTATCAACTTCTCCAAAGACC 621
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 GGTTCCTGTGAAGAAGTCCGAGAGATCTCTGAACCTGGACCGTATCAACTCTCTCCAAAGAC 597

QY 622 ATCACTTTCGAGATCAACAAGATCCCAAGACCGGTTGTGATCACTCCGACTCCCGACTCCGACAC 681
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 ATCGACTTCAAG-----CTGATGAACCTGTACCGACACACACC 633

QY 682 ATCAACATGTGGATCCGCTGACTTCTACATCTTCCCAAGAGGTTGCGGTAAGGACATC 741
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 AAGTGTGTTGATCAAGGACTTCAACATCTTCTGGTGTGAGCTGAACCCCGGAGGTC 693

QY 742 AACATCTGTTTCAACTCTCTGAGTACACCAAGCTGCTCAAGGACTACTGGGTAAACGAC 801
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 694 TCCCTCCCTGTACTGTGATCCAGTCTCTCCCAACACACCTGAAGGACTTCTGGGGAACCCCA 753

QY 802 CTGAGATACAACAGGAGTACTATAGTCAACATCGACTACTTGAACAGATACATGTAC 861
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 754 CTGGGTTACGACACCCAGTACTACTCTGTTTCAACCAAGGATATGAGAACATCTACATCAAG 813

QY 862 GCCAACTCCA---GACAGATCGCTTCAACACCAAGAGTAAACAAACAGCACTTCAACAGAG 918
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 814 TACTTCTCCAGGCTCTCCATGGTGGTGAACCGCCCTCGTACCAACTTCAACACAGCGCGCC 973
```



```
Query Match      12.3%; Score 168.8; DB 10; Length 1338;
Best Local Similarity 54.8%; Pred. No. 5.6e-33;
Matches 447; Conservative 0; Mismatches 327; Indels 42; Gaps 4;

QY 64 ATCATCAAGAGTACTTCAACAACTCAACGACTCCCAAGATCCTGTCCCTCGAGAACCGT 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27 ACCTTCACTGAATACATCAAGACATCATCAATACCTCCATCTGAACCTGGCTACGAA 86

QY 124 AAGAACACCTTGGTGCACCTCCGCTTACACGCGGAGGTCTCGAGGAGGTGACGTC 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 87 TCCAATCACCTGATCGACCTGTCTCGCTACGCTTCCAAATCAACATCGGTTCTAAAGTT 146

QY 184 CAGCTGAACCAATCTCCCATTCGACTCAAGCTGGTCTCCGCTGAGGACAGAGGT 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 147 AACTTCGATCCGATCGNCAAGATCAGATCCAGCTG-----TTCAATCTGGAATCTCC 200

QY 244 AAGTCACTGCTACCCAGAACGAGACATCGTCTCAACATCCATGATGACGTCCTTCCTCC 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 201 AAAATCGAAGTTATCTCGTGAAGATGCTATCGTATACAACTCTATGTACGAAAACCTCTCC 260

QY 304 ATCTCCTCTCTGGATCAGATCAACAAGTGGTCTCCAACTGCC-----AGGTTAC 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 261 ACCTCCTCTCTGGATCGGTATCCCGAATACTTCAACTCCATCTCTCTGAACATGAATAC 320

QY 355 ACCATCATGACTCCGCTCAGAAACAACTCCGCTTGGTCTCCATCGGTATCATCTCCAACTTC 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 321 ACCATCATCAACTGCTATGGAACAAATCTCTGGTTGGAAGTATCTCTGAACATACGGTGAA 380

QY 415 CTGCTCTTCAACCTGAGCAGACGAGGACTCCGAGCAGTCCATCACTTCTTCCATACGAC 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 ATCATCTGGACTCTGAGGACATCAGGAATCAACACGCGTGTGTATTAACAACTACTCT 440

QY 475 ATCTCCCAACACGCTCTCTGTTTAAACAA-----GTGTTTCTTCTGCTCACTCCACCAACAC 531
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 441 CAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTTCTGTACCATCACCAACAT 500

QY 532 ATGATGGGTAAATGATCTACATCAACGCTGAGTACGATGATGACACACATCAAGTCAAG 591
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 CGTCTGAATCACTCAAAATCTACATCAACGCGCTGTGATGACGACGAAACCGATCTCC 560

QY 592 GAGTTGACCGGTATCACTCTCCAGACATCACTCTCGATCAACAGATCCACAGAC 651
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 AATCTGGTAACTCCAGCTTCTAATACATCATGTTCAACTGGACGGTGTCTGATGAC 620

QY 552 ACCGGTCTGATCACTCCGACTCCGACATCAACATGATGATGATGATGATGATGATGATG 711
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 621 ACTC-----ACCGCTACATCTGGATCAAAATCACTCAATCTG 656

QY 712 TTCGCCAAGGAGTGGAGTAAAGGACATCAACATCCTCTTCAACTCTTCCGAGTACACC 771
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 657 TTCGACAAAGAACTGACGAAAGAAATCAAGACCTGTACGACACACGATCCAAATCT 716

QY 772 AACGCTCCTAAGGACTACTGGGTAACGACCTGAGATACAAAGAGGAGTACTACATGTC 831
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 717 GGTATCCTGAAGACTTCTGGGTGACTACTCTGCAGTACGACAAACCGTACTACATGCTG 776

QY 832 AACATGACTACTTGAACATACATGATAGCGCAAC 867
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 777 AATCTGTACGTCGGAACAAATACGTTGAGCTAAC 812

RESULT 6
US-09-350-756-4
; Sequence 4; Application US/09350756
; Patent No. US2002034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith

; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP R11D 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; EARLIER FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 4
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-4

Query Match      11.8%; Score 162.4; DB 10; Length 1400;
Best Local Similarity 54.4%; Pred. No. 2.3e-31;
Matches 468; Conservative 0; Mismatches 341; Indels 51; Gaps 5;

QY 16 ATCCCATTCACATCTTCTCTACACCAACAACCTCCCTGTTTGAAGGACATCATCAACGAG 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 ATTCTTTTAAAGTTTCTTCTTATACAGATGATAAAATTTTAAATTTCTTCTTCAACAG 126

QY 76 TACTTCAACAACTCAACGACTCCAAAGATCCTGTCCCTGCAGAACCGTAAGAACACCTTG 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 TTCTTCAAGAGAAATTAAGTCTTCTCCGTTTAAACATGAGATCAAGAATGATAAATAC 186

QY 136 GTGACACCTCCGGTTACACGCCGAGGCTCCGAGGAGGTTGAGTCCAGTGAAACCCA 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 GTGACACCTCCGGTTACGACTCCAAATATCAACATTAACGGTGGAGCTGTACAAAGTACC 246

QY 196 ATCTTCCCATTCGACTTCAAGCTGGTCTCTCCGTTGAGGACGAGGTAAGTCACTGTC 255
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 ACTAACAACCAACATTC-----GGTATCTACAAGCAAGCTTACTGAGCTGAACATC 300

QY 256 ACCGACGACGAGAAATCTGTCTACAACTCCATGTAGGAGTCTTCTTCCATCTCTCTTGG 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 TCTCAAAACGACTACATTTATCTACGACAAAGTACAAAGAACTTCTTATTTCTTCTGG 360

QY 316 ATCAGAAAT-----CAACAAGTGGTCTCCAACTTCCAGGTTACACCAAGATCATC 363
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GTCAGGATTCCTAACTACGACACAAAGTCTGTCAGTTTAAACAGAGTACACTATCATC 420

QY 364 GACTCCGTCA---AGAACAACTCCGTTGGTCCATCGGTATCATCTCCAACTTCTGTC 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 AACTGTATGAGACAAACAACCTCCGTTGGAAAGTCTCTCTTAAACCAACAGAGATCAT 480

QY 421 TTCACTGAGGACGACGAGGACTCCGAGCAGTCCATCAACTTCTCTTACGACATCTCC 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 TGGACCTTGAACACACGAGGATTAACCAAAAGTTAGCAATCACTACGCTAACGCA 540

QY 481 AACACGCTCCTGTTA---CAACAAGTGGTCTTGGTCCAGCTCACCACCAACATGATG 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 AACGGTATTTCTGACTACATCAACAAGTGGATTTTGGTCACTATCACTAACGACAGATTA 600

QY 538 GGTAACTGAAATCTACATCAACGCTAGCTGACGACACCATCAAGTCAAGGAGTTG 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 GGTGACTCTAAGCTTTACATTAACGGTAACTTAATGCGACCAAAAGTCCATTTTAAACTTA 660

QY 598 ACCGGTATCAACTTCTCCAAAGACCATCACTCTCGAGATCAACAAGATCCGACACACCGGT 657
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 GGTAACTTCAACGTTCTGCAACATCTTATTCAGATCGTTTAACTGCAAGTTACACGCA 720

QY 658 CTGATCACTCCGACTCCGACAAACATCAACATGTGGATCCGTTGATCTTCTTCTGCGC 717
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 T-----ACATTTGGCATTTAGATCTTCAACATTTTCGAC 753

QY 718 AAGGAGTTGACGAGTAAAGGACATCAACATCTCTGTTCACTCTTTCAGTACACCAACGTC 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 754 AAGAGTTAGACGAGACCGAGATTCAAATTTTATACAGCAACGAACTTAACACCAATTT 813

QY 778 GTCAAGGACTACTGGGTAAACGACTGAGATCAACAAGGAGTACTACATGTTGTTCAACATC 837
```

Db 814 TTGAAGGACTTCTGGGGTAAGTACTTCTTTACGACAGGAATACTACTATTATTAACGTTG 873

QY 838 GACTACTTGAACAGATACAT 857

Db 874 TTAAGCCAAACAACATTCAT 893

## RESULT 7

US-08-981-087A-6  
; Sequence 6, Application US/08981087A  
; Patent No. US20020081304A1  
; GENERAL INFORMATION:  
; APPLICANT: Elmore, Michael J.  
; APPLICANT: Nauchline, Margaret L.  
; APPLICANT: Minton, Nigel P.  
; APPLICANT: Pasechnik, Vladimir A.  
; APPLICANT: Titball, Richard W.  
; TITLE OF INVENTION: TYPE F BOTULINUM TOXIN AND USE THEREOF  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHVE P.C.  
; STREET: 1100 No. US20020081304A1th Glebe Rd. 8th floor  
; CITY: Arlington  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22201-4741  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/981.087A  
; FILING DATE: 27-MAY-1998  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/GB96/01409  
; FILING DATE: 12-JUN-1996

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 124-688  
; FILING DATE: 12-JUN-1995  
; APPLICATION NUMBER: GB 9511909.5  
; FILING DATE: 12-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crawford, Arthur R.  
; REGISTRATION NUMBER: 25,327  
; REFERENCE/DOCKET NUMBER: 124-688  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 703-816-4000  
; TELEFAX: 703-816-4100  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1313 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-981-087A-6

Query Match 11.1%; Score 151.6; DB 8; Length 1313;  
Best Local Similarity 53.3%; Fred. No. 1.1e-28;  
Matches 465; Conservative 0; Mismatches 359; Indels 48; Gaps 5;

QY 32 TCTCTACCAACCAACTCCCTGTTGAAGGACATCATCAACGAGTACTTCAACAACATCA 91

Db 10 TGTCTTACACTAGGACAAAATCCCTGATCTTCAACAACACTGTACAAAAATCA 69

QY 92 ACAGCTCCAGATCCCTGCTGCGAGACCGTAGAACAACCTTGGTGGACACCTCCGGTT 151

Db 70 AAGACAACCTCTATCCCTGGGACATCGCTTAGCAAAACAACAAATTCGACATCTCTGGCT 129

QY 152 ACAACGCGGAGGCTCCGAGGAGGCTGAGCTCCAGCTGAACCAATCTCCCATTCGACT 211

Db 130 ATGGTTCTACATCTATACACGGTGGAGCTGAGCTACTACTCTA-----CTAACGCA 183

QY 212 TCAGCTGGGTCTCTCCGGTGAGGACAGAGGTAAAGTCACTCGTCAACCAGAACGAGACA 271

Db 184 ACCAGTTCCGTATCTCTTCTTAACCCGCTCTGAAGTAAACATCGCTCAGAACAGAGACA 243

QY 272 TCGTCTACAACTCCATGTACGAGTCCCTTCCATCTCTTCTCTGATCAGAACTCAACAAGT 331

Db 244 TCACTACACGGTCGTACGAGAACTTCTATCTCTTCTTCTGGTTCGTATCCGAAAT 303

QY 332 GGGTCTCCAACTT-----GCCAGGTACACCATCATCGACTCCGGT----AAGAACA 379

Db 304 ACTTCACAAAAGTTAAAGTGAACAAACGAATACACTATCATCGACTGCATCCGTAACAACA 363

QY 380 ACTCGGTTGGTTCGATCGGTATCATCTCCCACTTCCTCCGCTTCCACCTGAACGAGAACG 439

Db 364 ACTCTGGTTGGAAATCTCTCTGAACTACACAAAATCATCTGGACTCTTCAGAGACACTG 423

QY 440 AGGACTCCGAGCAGTCCATCAACTTCTCCACGACATCTCCAAACAAGGCTCTCGGTAC- 498

Db 424 CTGGTAACAACGAGAACTGGTTTCAACTACACTCAGATGATCTCTATCTGAACTACA 483

QY 499 --AACAAGTGGTCTCTGTCACCGTCAACCAACATGATGGGTAACTGAAGATCTACA 556

Db 484 TTAATAAATGGATCTCTGTTACTATCACTAACAACCGCTCTGGGTAACTCTCGTATCTACA 543

QY 557 TCACGGTAACTGATGACACACCATCAAGGTCAAGGTCAAGGAGTTGACCGGTATCAACTCTCCA 616

Db 544 TCAACGGTAACTGATGATGAAAAATCTATCTTAACCTGGGTGACATCCAGCTTCTG 603

QY 617 AGACCATCACTTCGAGATCAACAAGATCCAGACACCGGCTGATCACTCCGACTCCG 676

Db 604 ACAACATCTGTTCAAAATCTTGGTTGCAACGACACCGCTT----- 645

QY 677 ACAACATCAACATGTGGATCGGTGACTCTTACATCTTCGCAAGGAGTTGGACGGTAAGS 736

Db 646 -----ACGTTGGTATCGTTACTTCAAGTTTTCGACACTGAACTGGTAAACTG 696

QY 737 ACATCAACATCTCTTCACTCTTCACTGAGTACACCAACGCTCGTCAAGGAGTACTGGGTA 796

Db 697 AAATCGAAACTCTGTACTCTCAACGACCGGACCGCTATCTCCTGAAAGACTCTGGGGTA 756

QY 797 AGGACTGAGATACAAAGGAGTACTACATGTTGTTCAACATCGACTTCTGAAACAGATACA 856

Db 757 ACTACTGCTGTACAAACAAAGTTACTTACCTGCTGAACTGCTCCGAGCTGACAAATCTA 816

QY 857 TGTAGCCAACTCCAGACAGATCGTCTTCAAC 888

Db 817 TCACCTCAGAACTCTAACTTCTCTGAACATCAAC 848

## RESULT 8

US-09-350-756-10  
; Sequence 10, Application US/09350756  
; Patent No. US20020034521A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases  
; APPLICANT: John S. Lee  
; APPLICANT: Peter Pushko  
; APPLICANT: Michael D. Parker  
; APPLICANT: Jonathan F. Smith  
; APPLICANT: Mark T. Dertzbaugh  
; APPLICANT: Leonard Smith  
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine  
; FILE REFERENCE: 003/124/SAP RIID 98-21  
; CURRENT APPLICATION NUMBER: US/09/350.756  
; CURRENT FILING DATE: 1999-07-09  
; EARLIER APPLICATION NUMBER: US 60/092.416  
; EARLIER FILING DATE: 1998-07-10  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: Apple Macintosh Microsoft Word 6.0  
; SEQ ID NO 10  
; LENGTH: 702  
; TYPE: DNA

```
; ORGANISM: Clostridium botulinum
; FEATURE:
US-09-350-756-10

Query Match      9.2%; Score 126.6; DB 10; Length 702;
Best Local Similarity 53.6%; Pred. No. 1.5e-22;
Matches 382; Conservative 0; Mismatches 289; Indels 42; Gaps 4;

QY 54 ATCATACAGAGTACTTCAACACATCAACGACTCCCAAGATCCTGTCCCTGCAGACCGT 123
DB 19 ACCCTTCTGTAATACATCAAGACATCAATACCTCCCTGAACCTGCCTACGAA 78

QY 124 AAGAACACCTTGGTCGACACCTCCGTTTACACGCGAGGTCTCCGAGAGGGTGACGTC 183
DB 79 TCCAATCAGCTGATCGACCTGTCTCGCTACGCTTCCAAAATCAACATCGTTCTRAAGTT 138

QY 184 CAGCTGAACCAATCTTCCATTCGACTTCAAGCTGGTTCCTCCGGTAGGACAGAGT 243
DB 139 AACTTCGATCGATCGACGACAAAGATCAGATCCAGCTG-----TTCAATCTGGAATCTTCC 192

QY 244 AAGGTCTATCGTCAACCCAGAACGATCGTCTACAACCTCCATGTACGAGTCCCTTCTCC 303
DB 193 AAAATCGAAGTTATCTCGAAGATGCTATCGTATACAACCTCTATGTACGAAAACITCTCC 252

QY 304 ATCTCCCTTCGGATCAGAAATCAACAGTGGTCTCAACTTGGC-----AGTTTAC 354
DB 253 ACCTCTCTCTGGATCCGTATCCCGAAATCTTCAACTCCATCTCTGAAACATGAATAC 312

QY 355 ACCATCATCGACTCCGTCAGAACAACTCCGGTGTGGTCCATCGTATCTCCCAACTTC 414
DB 313 ACCATCATCAACTCGATGGAAACAAATCTGTGGTGGAAAGTATCTCTGAACTACGGTGA 372

QY 415 CTGGTCTTCACCTGAGAGAGAGAGGACTCCGAGGAGTCCATCAACTTCTCCTACGAC 474
DB 373 ATCATCTGGAATCTGCGAGGACTTCAGGAATCAACAGACGGTGTGTATTAATAACTCT 432

QY 475 ATCTCCAAACAGCTCTCTGGTTTACAA--GTGGTCTTCTGTCACCGTFCACCAACAC 531
DB 433 CAGATGATCAACATCTCTGACTACATCAATCGTGGTGTGATCTCTGTTACCATCAACAT 492

QY 532 ATGATGGTACATGAGATCTACATCAACGGTAAAGTATGATGACACCATCAAGGTCAAG 591
DB 493 CGTCTGATTAACCTCAAAATCTACATCAACGGCGCTGTGATGACGAGAAACGATCTCC 552

QY 592 GAGTGTACCGGTATCAACTCTCCAGACATCACTTCGAGATCAACAGATCCGAC 651
DB 553 AATCTGGTACATCCAGGTTCTTAATACATCATGTTCAACTGGACGGTGTCTGTGAC 612

QY 652 ACCGCTGTGATCACTCCGACTCCGACATCAACATGATGATGATGATGATGATGATGAT 711
DB 613 ACTC-----ACCGCTACATCTGGATCAATATCTCAATCTG 648

QY 712 TTCGCAAGAGAGTGGAGGTAAGGACATCAACATCCTGTTCACCTCTTCA 764
DB 649 TTCGCAAGAGAGTGGAGGTAAGGACATCAACATCCTGTTCACCTCTTCA 701

RESULT 9
US-09-350-756-2
; Sequence 2, Application US/09350756
; Patent No. US20020034521A1
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09

; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 2
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-2

Query Match      8.8%; Score 120.4; DB 10; Length 1341;
Best Local Similarity 50.7%; Pred. No. 6.9e-21;
Matches 412; Conservative 0; Mismatches 346; Indels 54; Gaps 3;

QY 54 ATCATCAACGAGTACTTCAACACATCAACGACTCCCAAGATCCTGTCCCTGCAGACCGT 123
DB 10 ATGCCCAACAAATACAAATCCGAAATCCCTGAACAATATCATCTGAACCTGGTTACAAA 69

QY 124 AAGAACACCTTGGTCGACACCTCCGTTTACACGCGAGGTCTCCGAGAGGGTGACGTC 183
DB 70 GACAACAATCTGATCGATCTGTCTGGTTACGGTGCTAAAGTTGAGTATACGACGCTGTT 129

QY 184 CAGCTGAACCAATCTTCCATTCGACTTCAAGCTGGTTCCTCCGGTAGGACAGAGT 243
DB 130 GAAGTGAATGACAAAGAACCAAGTTCAAACTGA-----CCTCTTCCGCTAACTCT 177

QY 244 AAGGTCTATCGTCAACCCAGAACGATCGTCTACAACCTCCATGTACGAGTCCCTTCTCC 303
DB 178 AAGTTCGGTGTACTCAGATCAGAACATCATCTTCAACTCCGATTCCTCGGATTCCTGAC 237

QY 304 ATCTCCCTTCGGATCAGAAATCAACAGTGT-----GGTCTCCCAACTTG 345
DB 238 GTTTCCTTCTGGATCTCGTATCCGAAATACAGAACGACGGTATCCAGAAATACATCCAC 297

QY 346 CCAGGTACACCATCATCGCTCCGTCAGAACAACTCCGGTGTGGTCCATCGTATCATC 405
DB 298 AATGAATACACCATCATCACTCATGATGAGAAATAACTCTGGTTGGAAGATCTCCATCCG 357

QY 406 TCCAACTTCCTGGTCTTCCACCTGAGAGAGAGAGAGATCCGAGCAGTCCATCAACTTC 465
DB 358 GGTAAAGCTATCATCTGGACTCTGATGATATCAACGGTAAAGACCAAACTCTGTATTCTTC 417

QY 466 TCCTACGACATCTCCAAACAGCTCCCTGGTTTACAACAGTGGTTCCTGTCACCGTCCAC 525
DB 418 GAATACACATCCGTTGAAGACATCTCTGAATACATCAATCGCTGGTTCCTGTTACCATC 477

QY 526 AACAAATGATGGTAAACATGAAGATCTACATCAACGGTAAAGTGTATGATGACACCATCAAG 585
DB 478 ACCAATAACCTGAAACAATGCTAAAATCTACATCAACGGTAAACTGGAATCTAATACCGAC 537

QY 586 GTCAAGAGATGTACCGGTATCAACTTCTCAAGACCATCACTTCCTGAGATCAACAGATC 645
DB 538 ATCAAGACATCCGTTGAAGTATCGTTAAAGGTGAATCATCTTCAAACTGGACGGTGC 597

QY 646 CCAGACACCGGTCTGTGATCACTCCGACTCCGACACATCAACATGTGGATGGTGCATCTTC 705
DB 598 ATCGATCGTACCCAGTTTCACT-----TGATGAATACTTC 633

QY 706 TACATCTTCGCAAGAGAGTGGAGGTAAGGACATCAACATCCTGTTCACCTCTTGCAG 765
DB 634 TCCATCTTCAACACCGAACTGTCTCAGTCCCAATATCGAAGACGGTACAGATCTCCAGT 693

QY 766 TACACCAACGTCGTCAGAGGACTACTGGGGTAAAGGACTGTGATGATCAACAAAGAGTACTAC 825
DB 694 TACTCCGAAATACCTGAAAGACTTCTGGGGTAAATCCCGCTGATGTACAAACAAGAACTACTAT 753

QY 826 ATGTCAACATCGACTACTTGTGAACAGATACAT 857
DB 754 ATGTTCAATGCTGTTAAACAAGAACTCTTTACAT 785
```









```

; TYPE: DNA
; ORGANISM: Toxoplasma gondii
; US-09-216-393-342

Query Match      6.6%; Score 90.2; DB 10; Length 867;
Best Local Similarity 51.1%; Pred. No. 2e-13;
Matches 212; Conservative 0; Mismatches 203; Indels 0; Gaps

QY  734 AGACATCAACATCCTGTTCAACTCCTTCAGTACACCAACGTCGTCAGGACTACTGGG 793
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db   451 AGTTCATCCACAAGTAGTAGAGTACACGACCGAGCATCAACTACCCTACTACCCTACC 510
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY  794 GTAACGACCTGAGATACACAAGGAGTAGTACATGGTCAACATCGACTACTTTGAACAGAT 853
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db   511 ACCACTACGACTACTACTACAACTACGACACCAACAACAACCTACACAACCAACAACAACT 570
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY  854 ACATGTATGCCCACTCCAGACAGATGCTTTCAACACCAAGAGTCAACACGACGACTTCA 913
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db   571 ACACCAACACACGACACACACCCACAACTACCAACAAACGACACACACACACCA 630
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY  914 ACGAGGGTTACAGAGTCATCATCAAGCGTATCAGAGGTAAACACCACGACACGAGTCA 973
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db   631 ACAACTACACCAACAACGACACCAACCAACCAAGCCAACTACACACGACATCTACGACA 690
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY  974 GAGGTGTGACATCCGTGTTCTTCGACATGACTATCAACAACAAGCCCTACAACTGTTCA 1033
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db   691 ACCACTACGACTACCACACTACTACTACCCACTACACAAACGACAAACCCGGAACCA 750
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY  1034 TGAAGAACGAGACCATGTATCGCGGACCAACCACTCCACCGGAGGACATCTACGCCATCGGTC 1093
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db   751 ACAACTACACAACAACCCAGGAACCAACCAACCACTACAAAGCAACAACGACGACTACGACA 810
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY  1094 TCGGTGAGCAGACCAAGGACATCAACGCAACACTCATCTTCAGATCCAGCCAAAT 1148
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db   811 ACTACACGACTACGACACCCNCTACGACGACATCCACCACCACTACCCCTCGATT 865
      || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Search completed: November 7, 2002, 14:41:50
Job time : 68 secs

```

Search completed: November 7, 2002, 14:41:50  
Job time : 68 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:21:30 ; Search time 3155 Seconds  
(without alignments)  
10925.626 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371

Sequence: 1 gaattcacgagaccatccc.....tctccgagtaataaggaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_NA\_Main.\*  
1: /cgn2\_6/ptodata/1/pna/pctus\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US06\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US082\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US083\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US084\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US087\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US088\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US089\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US090\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US091\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US092\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US094\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US096B\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US096C\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US096D\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US096E\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US097A\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US097B\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US097C\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US098A\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US098B\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US098C\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US099A\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US099B\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US099C\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US099D\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US100A\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US100B\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US101A\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US101B\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US102A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US102B\_COMB.seq.\*

44: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US6002\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US6003\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US6004\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US6008\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US6009\_COMB.seq.\*  
54: /cgn2\_6/ptodata/1/pna/US6010\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US6014\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US6015\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US6016\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/US6017\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US6018\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US6019\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US6020\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US6021\_COMB.seq.\*  
66: /cgn2\_6/ptodata/1/pna/US6022\_COMB.seq.\*  
67: /cgn2\_6/ptodata/1/pna/US6023\_COMB.seq.\*  
68: /cgn2\_6/ptodata/1/pna/US6024\_COMB.seq.\*  
69: /cgn2\_6/ptodata/1/pna/US6025\_COMB.seq.\*  
70: /cgn2\_6/ptodata/1/pna/US6026\_COMB.seq.\*  
71: /cgn2\_6/ptodata/1/pna/US6027\_COMB.seq.\*  
72: /cgn2\_6/ptodata/1/pna/US6028\_COMB.seq.\*  
73: /cgn2\_6/ptodata/1/pna/US6029\_COMB.seq.\*  
74: /cgn2\_6/ptodata/1/pna/US6030\_COMB.seq.\*  
75: /cgn2\_6/ptodata/1/pna/US6031\_COMB.seq.\*  
76: /cgn2\_6/ptodata/1/pna/US6032\_COMB.seq.\*  
77: /cgn2\_6/ptodata/1/pna/US6033\_COMB.seq.\*  
78: /cgn2\_6/ptodata/1/pna/US6034\_COMB.seq.\*  
79: /cgn2\_6/ptodata/1/pna/US6035\_COMB.seq.\*  
80: /cgn2\_6/ptodata/1/pna/US6036\_COMB.seq.\*  
81: /cgn2\_6/ptodata/1/pna/US6037\_COMB.seq.\*  
82: /cgn2\_6/ptodata/1/pna/US6038\_COMB.seq.\*  
83: /cgn2\_6/ptodata/1/pna/US6039\_COMB.seq.\*  
84: /cgn2\_6/ptodata/1/pna/US6040\_COMB.seq.\*  
85: /cgn2\_6/ptodata/1/pna/US6041\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1371	100.0	1371	23	US-09-611-419A-9
2	1371	100.0	1371	34	US-09-910-186A-9
3	1366.2	99.6	1371	17	US-09-350-756-3
4	685.2	48.5	3876	1	PCT-US97-15394-59
5	685.2	48.5	3876	11	US-08-704-159-59
6	685.2	48.5	3876	42	US-10-205-516-5
7	685.2	48.5	3876	43	US-10-271-012-59
8	665.2	48.5	3950	13	US-08-954-302-2
9	663.2	48.4	3906	42	US-10-205-516-19
10	662	48.3	3876	29	US-09-730-237-4
11	662	48.3	3876	39	US-10-051-952-4
12	649.8	47.4	1502	1	PCT-US97-15394-61
13	649.8	47.4	1502	11	US-08-704-159-61
14	649.8	47.4	1502	43	US-10-271-012-61
15	420.2	30.6	1374	23	US-09-611-419A-11
16	417	30.4	1374	34	US-09-910-186A-11
17	217.4	15.9	1347	17	US-09-350-756-6
18	217.4	15.9	1368	23	US-09-611-419A-17
19	217.4	15.9	1368	34	US-09-910-186A-17
20	199.8	14.6	1317	23	US-09-611-419A-15
21	199.8	14.6	1317	34	US-09-910-186A-15

```
22 198.2 14.5 1317 17 US-09-350-756-5 Sequence 5, Appl1
23 169.6 12.4 1351 1 PCT-US97-15394-35 Sequence 35, Appl
24 169.6 12.4 1351 11 US-08-704-159-35 Sequence 35, Appl
25 169.6 12.4 1351 43 US-10-271-012-35 Sequence 35, Appl
26 169.4 12.4 1323 23 US-09-611-419A-3 Sequence 3, Appl1
27 169.4 12.4 1323 34 US-09-910-186A-3 Sequence 3, Appl1
28 169.4 12.4 1402 1 PCT-US97-15394-25 Sequence 25, Appl
29 169.4 12.4 1402 8 US-08-405-496-25 Sequence 25, Appl
30 169.4 12.4 1402 8 US-08-422-711-25 Sequence 25, Appl
31 169.4 12.4 1402 8 US-08-480-604-25 Sequence 25, Appl
32 169.4 12.4 1402 11 US-08-704-159-25 Sequence 25, Appl
33 169.4 12.4 1402 14 US-09-084-517-25 Sequence 25, Appl
34 169.4 12.4 1402 43 US-10-271-012-25 Sequence 25, Appl
35 168.8 12.3 1326 23 US-09-611-419A-5 Sequence 5, Appl1
36 168.8 12.3 1326 34 US-09-910-186A-5 Sequence 5, Appl
37 168.8 12.3 1330 1 PCT-US97-15394-22 Sequence 22, Appl
38 168.8 12.3 1330 8 US-08-405-496-22 Sequence 22, Appl
39 168.8 12.3 1330 8 US-08-422-711-22 Sequence 22, Appl
40 168.8 12.3 1330 8 US-08-480-604-22 Sequence 22, Appl
41 168.8 12.3 1330 11 US-08-704-159-22 Sequence 22, Appl
42 168.8 12.3 1330 14 US-09-084-517-22 Sequence 22, Appl
43 168.8 12.3 1330 43 US-10-271-012-22 Sequence 22, Appl
44 168.8 12.3 1332 23 US-09-611-419A-1 Sequence 1, Appl1
45 168.8 12.3 1332 34 US-09-910-186A-1 Sequence 1, Appl1
```

## ALIGNMENTS

## RESULT 1

```
; Sequence 9, Application US/09611419A
; GENERAL INFORMATION:
; APPLICANT: Smith, Leonard A.
; APPLICANT: Byrne, Michael P.
; APPLICANT: Middlebrook, John L.
; APPLICANT: Lapenotiere, Hugh
; APPLICANT: Clayton, Michael A.
; APPLICANT: Brown, Douglas R.
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; FILE REFERENCE: A33626 067252.0105
; CURRENT APPLICATION NUMBER: US/09/611.419A
; CURRENT FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/146,192
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on BONTA Hc
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)...(1359)
US-09-611-419A-9
```

```
Query Match 100.0%; Score 1371; DB 23; Length 1371;
Best Local Similarity 100.0%; Pred. No. 3.8e-279;
```

```
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAATTCAGGATGACATCCCATTCACATCTTCTCTACACCAACACTCCCTGTTGAAG 60
Db 1 GAATTCAGGATGACATCCCATTCACATCTTCTCTACACCAACACTCCCTGTTGAAG 60
Qy 61 GACATCATCAACGAGTACTTCCACACATCAAGACTTCCAGATCTCTGTCGAGAAC 120
Db 61 GACATCATCAACGAGTACTTCCACACATCAAGACTTCCAGATCTCTGTCGAGAAC 120
Qy 121 CGTAGAACACCTTGGTCGACACCTCCGGTTACACGCCAGGCTCCCGAGGAGGTGAC 180
Db 121 CGTAGAACACCTTGGTCGACACCTCCGGTTACACGCCAGGCTCCCGAGGAGGTGAC 180
Qy 181 GTCCAGCTGAACCCAAATCTTCCCATTCGACTTCAAGTGGGTCTCCCGGTGAGGACAG 240
Db 181 GTCCAGCTGAACCCAAATCTTCCCATTCGACTTCAAGTGGGTCTCCCGGTGAGGACAG 240
Qy 241 GGTAGGTCATCGTCACCCAGACGAGAAATCGTCTACAACTCCATGTACGAGTCTCTTC 300
Db 241 GGTAGGTCATCGTCACCCAGACGAGAAATCGTCTACAACTCCATGTACGAGTCTCTTC 300
Qy 301 TCCATCTCTTCTGATCAGAAATCAACAAGTGGGTCTCCAACTTGCAGGTTACACCATC 360
Db 301 TCCATCTCTTCTGATCAGAAATCAACAAGTGGGTCTCCAACTTGCAGGTTACACCATC 360
Qy 361 ATCGACTCCGTCAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTCTGGTC 420
Db 361 ATCGACTCCGTCAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTCTGGTC 420
Qy 421 TTCACCTTGAAGCAGACGAGGACTCCGAGCAGTCCCACTCAACTTCTCTACGACATCTCC 480
Db 421 TTCACCTTGAAGCAGACGAGGACTCCGAGCAGTCCCACTCAACTTCTCTACGACATCTCC 480
Qy 481 AACACGCTCTCTGTTTACACAAAGTGGTCTTCTGTCACCGTCACCAACACATGATGGGT 540
Db 481 AACACGCTCTCTGTTTACACAAAGTGGTCTTCTGTCACCGTCACCAACACATGATGGGT 540
Qy 541 AACATGAAGATCTACATCAACCGTAAAGTGTATGACACACCATCAAGGTCAAGGAGTGAAC 600
Db 541 AACATGAAGATCTACATCAACCGTAAAGTGTATGACACACCATCAAGGTCAAGGAGTGAAC 600
Qy 601 GGTATCAACTCTCCAGAACCATCACTTCGAGATCAACAGATCCAGACACCGGTCTG 660
Db 601 GGTATCAACTCTCCAGAACCATCACTTCGAGATCAACAGATCCAGACACCGGTCTG 660
Qy 661 ATCACTCCGACTCCGACAAACATCAACATGTGATCGGTCTTCTACATCTTCCGCCAAG 720
Db 661 ATCACTCCGACTCCGACAAACATCAACATGTGATCGGTCTTCTACATCTTCCGCCAAG 720
Qy 721 GAGTTGGACGATAGGACATCAACATCTCTCAACTCCTTTCAGTACACCAACGTCGTC 780
Db 721 GAGTTGGACGATAGGACATCAACATCTCTCAACTCCTTTCAGTACACCAACGTCGTC 780
Qy 781 AAGGACTACTGGGTAAAGGATCAACATCTCTCAACTCCTTTCAGTACACCAACGTCGTC 840
Db 781 AAGGACTACTGGGTAAAGGATCAACATCTCTCAACTCCTTTCAGTACACCAACGTCGTC 840
Qy 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTTCACACCAACGTCGTC 900
Db 841 TACTTGAACAGATACATGTACGCCAACTCCAGACAGATCGTCTTTCACACCAACGTCGTC 900
Qy 901 AACACGACTTCAACGAGGTTTACAGATCATCATCAGCGTATCAGAGGTATACACCAAC 960
Db 901 AACACGACTTCAACGAGGTTTACAGATCATCATCAGCGTATCAGAGGTATACACCAAC 960
Qy 961 GACACGAGTCCAGAGGTTGTTGACATCTCTTCTGACATGACTATCAACAAAGGCC 1020
Db 961 GACACGAGTCCAGAGGTTGTTGACATCTCTTCTGACATGACTATCAACAAAGGCC 1020
Qy 1021 TACAACCTGTTTATGAAGAACGAGACCATGTACGCCGACACCACTCCACCGAGGATC 1080
Db 1021 TACAACCTGTTTATGAAGAACGAGACCATGTACGCCGACACCACTCCACCGAGGATC 1080
```

```
QY 1081 TACGCCATCGGTCTGCGTACGAGACCAAGGACATCAACGAGCAACATCATCTTCCAGATC 1140
Db 1081 TACGCCATCGGTCTGCGTACGAGACCAAGGACATCAACGAGCAACATCATCTTCCAGATC 1140
QY 1141 CAGCCATGACACACTTACTACTAGCTTCCAGATCTTCCAGATCTTCAAGCTTCAACGGT 1200
Db 1141 CAGCCATGACACACTTACTACTAGCTTCCAGATCTTCCAGATCTTCAAGCTTCAACGGT 1200
QY 1201 GAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGATCTCGTCTGGGTGACTGG 1260
Db 1201 GAGAACATCTCCGGTATCTGTTCCATCGGTACCTACAGATCTCGTCTGGGTGACTGG 1260
QY 1261 TACAGACACACTTGGTTCCTCACTGCTCAACGAGGTAACCTACGCTTCTTGTGTGGAG 1320
Db 1261 TACAGACACACTTGGTTCCTCACTGCTCAACGAGGTAACCTACGCTTCTTGTGTGGAG 1320
QY 1321 TCCACTTCCACCCACTGGGATTCGCCAGTCTCCGAGTAAATAGGAATTC 1371
Db 1321 TCCACTTCCACCCACTGGGATTCGCCAGTCTCCGAGTAAATAGGAATTC 1371
```

## RESULT 2

```
US-09-910-186A-9
; Sequence 9, Application us/09910186A
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research & Materiel Command
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; FILE REFERENCE: A33626-A 067252.0107
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US/09/910,186A
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: PCT/US00/12890
; PRIOR FILING DATE: 2000-05-12
; PRIOR APPLICATION NUMBER: 09/611,419
; PRIOR FILING DATE: 2000-07-06
; PRIOR APPLICATION NUMBER: 60/133,865
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,866
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,867
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,868
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,869
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 60/133,873
; PRIOR FILING DATE: 1999-05-12
; PRIOR APPLICATION NUMBER: 08/123,975
; PRIOR FILING DATE: 1993-09-21
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum
; NAME/KEY: CDS
; LOCATION: (10)...(1359)
US-09-910-186A-9
```

```
Query Match 100.0%; Score 1371; DB 34; Length 1371;
Best Local Similarity 100.0%; Pred. No. 3.8e-279;
Matches 1371; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GAATTCAGATGACCATCCCATTCATCTTCTTACACCAACCACTCCCTGTTGAAG 60
Db 1 GAATTCAGATGACCATCCCATTCATCTTCTTACACCAACCACTCCCTGTTGAAG 60
QY 61 GACATCATCAGAGTACTTCAACACATCAACGACTCCAGATCTTCCCTTCGAGAAC 120
```

```
Db 61 GACATCATCAGAGTACTTCAACACATCAACGACTCCAGATCTTCCCTTCGAGAAC 120
QY 121 CGTAAGAACACCTTGGTTCGACACCTCCGGTTACAACGCCGAGGTTCTCGAGGAGGTTGAC 180
Db 121 CGTAAGAACACCTTGGTTCGACACCTCCGGTTACAACGCCGAGGTTCTCGAGGAGGTTGAC 180
QY 181 GTCCAGTGAACCAATCTTCCATTCGACTTCAAGCTGGGTTCTCCGGTTCGAGACAGA 240
Db 181 GTCCAGTGAACCAATCTTCCATTCGACTTCAAGCTGGGTTCTCCGGTTCGAGACAGA 240
QY 241 GGTAAAGTCTATCGTCCAGAACGAGAACATCGTCTACAACCTCCATGTCAGAGTCCCTTC 300
Db 241 GGTAAAGTCTATCGTCCAGAACGAGAACATCGTCTACAACCTCCATGTCAGAGTCCCTTC 300
QY 301 TCCATCTCTTCTTGGATCAGATCAACAAAGTGGGTTCCAACTTGCAGGTTTACACATC 360
Db 301 TCCATCTCTTCTTGGATCAGATCAACAAAGTGGGTTCCAACTTGCAGGTTTACACATC 360
QY 361 ATCGACTCCGTCGAAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCCTGGTC 420
Db 361 ATCGACTCCGTCGAAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCCCTGGTC 420
QY 421 TTCACCTGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTTCCAGACATCTCC 480
Db 421 TTCACCTGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTTCCAGACATCTCC 480
QY 481 AACACGCTCCTCGTTACACAAAGTGGTTCCTGTCACCGTCAACCAACAAATGATGGT 540
Db 481 AACACGCTCCTCGTTACACAAAGTGGTTCCTGTCACCGTCAACCAACAAATGATGGT 540
QY 541 AACATGAAGTCTTACATCAACGGTAAAGTGTGACACCATCAAGTCAAGGTTGACC 600
Db 541 AACATGAAGTCTTACATCAACGGTAAAGTGTGACACCATCAAGTCAAGGTTGACC 600
QY 601 GGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAAAGTCCCAACACCGGTTCTG 660
Db 601 GGTATCAACTTCTCCAGACCATCACTTCGAGATCAACAAAGTCCCAACACCGGTTCTG 660
QY 661 ATCACTCCGACTCCGACAACTCAACATGTTGGATCCGTGACTTCTACATCTTCGCCAAG 720
Db 661 ATCACTCCGACTCCGACAACTCAACATGTTGGATCCGTGACTTCTACATCTTCGCCAAG 720
QY 721 GAGTTGACGGTAAAGGACATCAACATCTTCAACTCTTCAAGTTCAGTACACCACTGCTC 780
Db 721 GAGTTGACGGTAAAGGACATCAACATCTTCAACTCTTCAAGTTCAGTACACCACTGCTC 780
QY 781 AAGGACTACTGGGTACGACCTGAGATACAACAAAGGAGTACTACATGTTCAACATCGAC 840
Db 781 AAGGACTACTGGGTACGACCTGAGATACAACAAAGGAGTACTACATGTTCAACATCGAC 840
QY 841 TACTTGAACAGATACATGTACGCCAATCCAGACAGATCGTCTTCAACACCAAGAGTAAAC 900
Db 841 TACTTGAACAGATACATGTACGCCAATCCAGACAGATCGTCTTCAACACCAAGAGTAAAC 900
QY 901 AACACGACTTCAACGAGGTTTCAAGATCATCATCAAGCGTATCAGAGTAAACACCAAC 960
Db 901 AACACGACTTCAACGAGGTTTCAAGATCATCATCAAGCGTATCAGAGTAAACACCAAC 960
QY 961 GACACAGAGTCAAGAGTGGTGTGATCTTCTTCAAGTTCAGATGATATCAACACCAAGGCC 1020
Db 961 GACACAGAGTCAAGAGTGGTGTGATCTTCTTCAAGTTCAGATGATATCAACACCAAGGCC 1020
QY 1021 TACAACTGTTTATGAGAACGAGACCATGTACGCCGACCAACCACTCCACCGAGGACATC 1080
Db 1021 TACAACTGTTTATGAGAACGAGACCATGTACGCCGACCAACCACTCCACCGAGGACATC 1080
QY 1081 TACGCCATCGGTCTGCGTACGAGACCAAGGACATCAACGACAAATCATCTTCCAGATC 1140
Db 1081 TACGCCATCGGTCTGCGTACGAGACCAAGGACATCAACGACAAATCATCTTCCAGATC 1140
QY 1141 CAGCAATGACACACTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGT 1200
Db 1141 CAGCAATGACACACTTACTACTAGCTTCCAGATCTTCAAGTCCCACTTCAACGGT 1200
```

```
QY 1201 GAGACATCTCCGGTATCTGTTCATCGGTACCTACAGATTCCGTCGTGGTGGTGAAGTGG 1260
Db 1201 GAGACATCTCCGGTATCTGTTCATCGGTACCTACAGATTCCGTCGTGGTGGTGAAGTGG 1260
QY 1261 TACAGACACAACACTACTTGGTCCCAACTGTCAAGCAGGTAACACTCGCTCCTTGTCTGGAG 1320
Db 1261 TACAGACACAACACTACTTGGTCCCAACTGTCAAGCAGGTAACACTCGCTCCTTGTCTGGAG 1320
QY 1321 TCCACTTCCACCCTCGGGGATTCGTCCTCCAGTCTCCGAGTAATAGGAATTC 1371
Db 1321 TCCACTTCCACCCTCGGGGATTCGTCCTCCAGTCTCCGAGTAATAGGAATTC 1371

RESULT 3
US-09-350-756-3
; Sequence 3, Application US/09350756
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP RIID 98-21
; CURRENT APPLICATION NUMBER: US/09/350.756
; EARLIER FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 3
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-3

Query Match 99.6%; Score 1366.2; DB 17; Length 1371;
Best Local Similarity 99.8%; Pred. No. 3.9e-278;
Matches 1368; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAATTCACGATCACCATCCCATTCACATCTCTCTCTACACCAACACTCCCTGTTGAAG 60
Db 1 GAATTCACGATCACCATCCCATTCACATCTCTCTCTACACCAACACTCCCTGTTGAAG 60
QY 61 GACATCATCAACGAGTACTTCAACAACATCAACGACTCCAGATCCCTGTCGCGAAG 120
Db 61 GACATCATCAACGAGTACTTCAACAACATCAACGACTCCAGATCCCTGTCGCGAAG 120
QY 121 CGTAAGACACCTTGGTCGACACCTCCGGTTACAGCGCGAGGTCTCCGAGGAGGGTGAC 180
Db 121 CGTAAGACACCTTGGTCGACACCTCCGGTTACAGCGCGAGGTCTCCGAGGAGGGTGAC 180
QY 181 GTCCAGCTGAACCCCAATCTCCCATTCGACTTCAAGTGGGTCTCCCTCGGTGAGGACAGA 240
Db 181 GTCCAGCTGAACCCCAATCTCCCATTCGACTTCAAGTGGGTCTCCCTCGGTGAGGACAGA 240
QY 241 GGTAGGTGATCGTCCACGAGACGAGAACATCGTCTACAACTCCATGTCAGTGCCTTC 300
Db 241 GGTAGGTGATCGTCCACGAGACGAGAACATCGTCTACAACTCCATGTCAGTGCCTTC 300
QY 301 TCCATCTCTCTCTGATCAGATCAACAGTGGGTCTCCACTTGCAGGTTACACCATC 360
Db 301 TCCATCTCTCTCTGATCAGATCAACAGTGGGTCTCCACTTGCAGGTTACACCATC 360
QY 361 ATCGACTCCGTCAAGAACAACTCCGGTTGGTTCATCGGTATCATCTCCAACTTCCTGGTC 420
Db 361 ATCGACTCCGTCAAGAACAACTCCGGTTGGTTCATCGGTATCATCTCCAACTTCCTGGTC 420
```

```
QY 421 TTCACCCCTGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCCTACGACATCTCC 480
Db 421 TTCACCCCTGAAGCAGACGAGGACTCCGAGCAGTCCATCAACTTCTCCTACGACATCTCC 480
QY 481 AACACGCTCCTCGTGTACACAACAGTGGTCTTCTCGTCAACCGTCCACCAACACATGATGGGT 540
Db 481 AACACGCTCCTCGTGTACACAACAGTGGTCTTCTCGTCAACCGTCCACCAACACATGATGGGT 540
QY 541 AACATGAAGATCTACATCAACCGTAAAGTAACTGATCGACACCATCAAGTCAAGGTCAGGAGTTC 600
Db 541 AACATGAAGATCTACATCAACCGTAAAGTAACTGATCGACACCATCAAGTCAAGGTCAGGAGTTC 600
QY 601 GGTATCAACTTCTCCAGACCATCACCTTCGAGATCAACAGATCCCGAGACACCGGTCTG 660
Db 601 GGTATCAACTTCTCCAGACCATCACCTTCGAGATCAACAGATCCCGAGACACCGGTCTG 660
QY 661 ATCACTCCGACTCCGACACATCAACATGTGGATCGGTGACTTCTACATCTTCGCGCAAG 720
Db 661 ATCACTCCGACTCCGACACATCAACATGTGGATCGGTGACTTCTACATCTTCGCGCAAG 720
QY 721 GAGTTGGAGCGTAAAGGACATCAACATCCTGTTCACCTTGCAGTACACCAACGCTCGTC 780
Db 721 GAGTTGGAGCGTAAAGGACATCAACATCCTGTTCACCTTGCAGTACACCAACGCTCGTC 780
QY 781 AAGGACTACTGGGTAGGACTCGATGATCAACAGGAGTACTACATGTTCAACATCGAC 840
Db 781 AAGGACTACTGGGTAGGACTCGATGATCAACAGGAGTACTACATGTTCAACATCGAC 840
QY 841 TACTTGAACAGATACATGTAGCCCAACTCCAGACAGATCGTCTTCAACACGACGCTAAC 900
Db 841 TACTTGAACAGATACATGTAGCCCAACTCCAGACAGATCGTCTTCAACACGACGCTAAC 900
QY 901 AACACGACTTCAACGAGGTTCAAGATCATCATCAAGCGTATCAGAGTAAACACCAAC 960
Db 901 AACACGACTTCAACGAGGTTCAAGATCATCATCAAGCGTATCAGAGTAAACACCAAC 960
QY 961 GACACGAGTCAGAGGTGGTACATCCTGTACTTCGACATGACTATCAACAACAAGGCC 1020
Db 961 GACACGAGTCAGAGGTGGTACATCCTGTACTTCGACATGACTATCAACAACAAGGCC 1020
QY 1021 TACAACCTGTTTATGAAGACGAGACCATGTACGCCGACCAACCACTCCACCGAGGACATC 1080
Db 1021 TACAACCTGTTTATGAAGACGAGACCATGTACGCCGACCAACCACTCCACCGAGGACATC 1080
QY 1081 TACGCCATCGGTCTCGGTGACGACCAAGGACATCAACGACACATCATCTTCCAGATC 1140
Db 1081 TACGCCATCGGTCTCGGTGACGACCAAGGACATCAACGACACATCATCTTCCAGATC 1140
QY 1141 CAGCCCAATGAACAACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200
Db 1141 CAGCCCAATGAACAACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGT 1200
QY 1201 GAGACATCTCGGTATCTGTTCATCGGTACCTACAGATTCGTCGTGGTGGTGAAGTGG 1260
Db 1201 GAGACATCTCGGTATCTGTTCATCGGTACCTACAGATTCGTCGTGGTGGTGAAGTGG 1260
QY 1261 TACAGACACAACACTACTTGGTTCCAACTGTCAAGCAGGTAACACTCGCTCCTTCTGGAG 1320
Db 1261 TACAGACACAACACTACTTGGTTCCAACTGTCAAGCAGGTAACACTCGCTCCTTCTGGAG 1320
QY 1321 TCCACTTCCACCCTCGGGGATTCGTCCTCCAGTCTCCGAGTAATAGGAATTC 1371
Db 1321 TCCACTTCCACCCTCGGGGATTCGTCCTCCAGTCTCCGAGTAATAGGAATTC 1371
```

```
RESULT 4
PCT-US97-15394-59
; Sequence 59, Application PC/TUS9715394
; GENERAL INFORMATION:
; APPLICANT: Williams, James A.
; APPLICANT: Thalley, Bruce S.
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium
; TITLE OF INVENTION: Botulinum Neurotoxin
```



NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/15394  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPD-02304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3876 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3873  
PCT-US97-15394-59

Query Match 48.5%; Score 665.2; DB 1; Length 3876;  
Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACATCCCATTCACATCTCTCTCCACCACTCCCTGTTGAAAGGACATCATCAAC 72  
DB 2527 ACAATACCCCTTTAATATTTTTCATATACATAATAATTTTATTAAGAGATATAATTAAT 2586

QY 73 GAGTACTCTCAACACATCAACGACTCCAGATCCCTGCTCCGAGAGGAGGTAAGGTGATC 132  
DB 2587 GAATATTCATTAATTAATGATTCATAAATTTGAGCCACAAACAGAAAATACT 2646

QY 133 TTGGTCGACACCTCCGGTTACAAACGCGGAGGCTCCGAGGAGGAGGTCAGCTCAGCTGAAC 192  
DB 2647 TTAGTGATACATCAGGATATAATGAGAAAGTGAGTGAGAGAGGCGATGTTTCAGCTTAAT 2706

QY 193 CCAATCTTCCATTCGACTTCAAGCTGGGTTCTCCGGTGAGGACAGAGGTAAGGTGATC 252  
DB 2707 CCAATATTCATTCGACTTCAATTAATGATGTTTCCAGGAGGAGTAAAGTTATA 2766

QY 253 GTACCCGAGAACGAGAACATCGTCTACAACTCCAGTACGAGCTCTTCCATCTCCCTTC 312  
DB 2767 GTAACCCAGAAATATTTGATATAATTTCTATGATGAAGTTTTCAGCAITTAGTTT 2826

QY 313 TGATCGATACATCAAGTGGGCTCCCACTCCGAGGTTACACCATCATCGACTCCGTC 372  
DB 2827 TGGATTAGATAATAATAATGGGTAAGTAAATTTACCTGGGATATATTAATGATAGTGT 2886

QY 373 AAGAACAACCTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTCTGGTCTTCCCTCGTGAAG 432  
DB 2887 AAAAAAATCACTGAGGTGGATATAGGTATTAATTAATTTTATAGTATTTTACTTTAA 2946

QY 433 CAGAACGAGACTCCGAGGAGTCCCACTCTCTCTAGACATCTCCACACACGCTCCT 492  
DB 2947 CAAATGAAGATAGTGAACAAAGATATAAATTTAGTTATGATATATCAATTAATGCTCCT 3006

QY 493 GGTACAAACAGTGGTCTCTCGTCCGCTCAGCAACACATGATGGTAAACATGAAGATC 552

DB 3007 GGATACAATAAATGGTTTTTTTCTAACTGTTTACTAACAAATATGATGGAAATATGAAGATT 3066

QY 553 TACATCAACGGTTAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACTTC 612

DB 3067 TATATAAATGGAAATTAATAGATATATAAAGTTAAAGTAACCTAAGTGAATTAATTTT 3126

QY 613 TCCAAGACCATCACCTTCGAGATCAACAAAGATCCAGACACCGGTCTGTATCACTCCGAC 672

DB 3127 AGCAAACTATAAACAATTTGAATAAATAAATTCAGATACCGGTTTGAATTTACTTCAGAT 3186

QY 673 TCCGACACATCAACATGTGGATCGTGACTTCTACATCTTCGCAAGGAGTTGGACGGT 732

DB 3187 TCTGATACATCAATATGTGGATAGAGATTTTATATATTTTGTAAAGATTAAGTGGT 3246

QY 733 AAGGACATCAACATCTCTTCAACTCTTGCAGTACACCACTGCTCAAGGAGTACTG 792

DB 3247 AAAGATATAATATATTTAATAGCTTGCATATACTAATGTGTGAAAGATTAATGG 3306

QY 793 GGTAAAGGCTGAGATACACAAAGGAGTACTACATGGTCAACATCGACTTCTGAACAGA 852

DB 3307 GGAATGATTAAGATATAATAAAGATATATATATGTTTATATAGATTTATTAATAGA 3366

QY 853 TACATGTACGCCAACTCCAGACAGATCGTCTTCAACACCAAGCTTAACAAACGACTTC 912

DB 3367 TATGTATGCGAACTCAGACAAATTTGTTTAAATACACGTAGAAATAATAATGACTTC 3426

QY 913 AACGAGGTTCAAGATCATCATCAAGCTATCAGAGGTATCAGAGGTACACCAACGACAGATC 972

DB 3427 AATGAGGATATAAATAATATAAAGAAATCAGAGGAAATCAAAATGATATAGTAGTA 3486

QY 973 AGAGTGGTGACATCTCTGACTTCGACATGACTATCAACAAACAAAGGCTTCAACCTGTC 1032

DB 3487 CGAGGAGGAGATATTTTATATTTTGTATGACATTAATAACAAAGCATATAATTTGTT 3546

QY 1033 ATGAAGACGAGACATGTACCCGACAAACACTCCACCGAGGACATCTAGCCATCGGT 1092

DB 3547 ATGAGATGAACATATGTATCGATATATCATAGTACTGAAGATATATATGCTATAGGT 3606

QY 1093 CTGCGTGACGACCAAGGACATCAACGACAACTATCTTCCAGATCCAGCCAAATGAAC 1152

DB 3607 TTAAGACAAACAAAGGATATAAATGATATATATTTCAATACAAACCAATGAAT 3666

QY 1153 AACATTTACTAGCTTCCAGATCTTCAAGTCCCACTTCAAGGTTGAGAACATCTCC 1212

DB 3667 AATACATTAATTAACGATCTCAATATTTAAATCAAAATTTAATGAGAAATATTTCT 3726

QY 1213 GGTATCTGTCCATCGTACAGATTCCTGCTGCTGGTGGTACTGGTACAGACACAAC 1272

DB 3727 GGAATATGTTCAATAGTACTTATCGTTTAGCTGGAGGTGATGCTATAGACAAAT 3786

QY 1273 TACTTGTTCCACTGCAAGCAGGTAACACTACGCTCTTGTGCTGGAGTCCACTTCCAC 1332

DB 3787 TATTTGTGCTACTGTGAAGCAAGAAATTTGCTTCTATTATTAGAAATCAACATCACT 3846

QY 1333 CACTGGGATTCGTCCACTCTCCGAGTAA 1362

DB 3847 CATTTGGGTTTGTCTACCTGTAAAGTGAATA 3876

RESULT 5  
US-08-704-159-59  
; Sequence 59, Application US/08704159  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; APPLICANT: Thallay, Bruce S.  
; TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
; TITLE OF INVENTION: Botulinum Neurotoxin  
; NUMBER OF SEQUENCES: 82  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco

STATE: California  
 COUNTRY: United States of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA: US/08/704,159  
 FILING DATE:  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ingolia, Diane E.  
 REGISTRATION NUMBER: 40, 027  
 REFERENCE/DOCKET NUMBER: OPD-02304  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 703-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 59:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3876 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..3873  
 US-08-704-159-59

Query Match 48.5%; Score 565.2; DB 11; Length 3876;  
 Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
 Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATCAACATCTCTCCACACCAACACTCCCTGTTGAAGGACATCATCAAC 72  
 DB 2527 ACATACCCCTTTAATATTTTTCATATCTACTAATATCTTTTAAAGATATATTAAT 2586  
 QY 73 GAGTACTTCAACACATCAACGACTCCAGATCCCTGTCGAGACCGTAAGAACCC 132  
 DB 2587 GAATATTTCAATATTAATGATTCATAAATTTTGAGCCCTACAAACAGAAAAATACT 2646  
 QY 133 TTGGTCGACACCTCCGTTTACAAACCCGAGGTCCTCGAGGAGGTCAGCTCCAGCTGAAC 192  
 DB 2647 TTAGTGGATACATCAGGATATATTCGAAAGTGTGAGAGCGGATGTCAGCTTAAT 2706  
 QY 193 CCAATCTTCCCATTCGACTTCAAGCTGGGTTCCCTCCGTCGAGGACAGAGGTAAGGTCATC 252  
 DB 2707 CCAATATTTCCATTTGACTTTAAATTTAGTAGTTCCAGGGGAGGATAGAGGTAAGTTATA 2766  
 QY 253 GTCACCCAGACGAGACATCTCTACACCTCCAGTACGAGTCCCTTCCTCATCTCCTTC 312  
 DB 2767 GTAACCCGAATGAATAATTTGTATATATCTATGATGAAGTTTATAGCTTAGTTT 2826  
 QY 313 TGGATCAGATCAACAAAGTGGTCTCCACTTGCAGAGTTACACATCATCGACTCCGTC 372  
 DB 2827 TGGATTAGTAATAAATAAGTGGTAAAGTAATTTACCTGGATATCTATATAATGATAGTTT 2886  
 QY 373 AAGAACTCCGTTGGTCCATCGGTATCTCCACTTCCCTGGTCTTCCACCTGAG 432  
 DB 2887 AAAATAACTCAGGTGGAGTAGTAGTATTTAGTAAATTTTAGTATTTACTTTAAA 2946  
 QY 433 CAGAAGCAGGACTCCGAGCAGTCCATCACTTCTCCTACGACATCTCCAAACAGCTCCT 492  
 DB 2947 CAAATGAAGATAGTAACAAAGTAAATTTTAGTTATGATATATCAAAATATGCTCCT 3006  
 QY 493 GGTACAAAGTGGTTCTTCGTCACCGCTCCACACACATGATGGTAACATGAAGATC 552  
 DB 3007 GGATCAATAAATGGTTTTTGTAACTGTACTAACAATATGATGGGAAATATGAAGATT 3066  
 QY 553 TACATCAACGGTAAAGTGTATCGACACCATCAAGGTCAGAGGTTACCGGTTATCAACTTC 612

DB 3067 TATATAAATGGAATAATTAATAGATACTATATAAAGTTAAAGAACTAACTGGAATTAATTTT 3126  
 QY 613 TCCAAGACCATCACCTTCGAGATCAACAAGATCCGACACACCGTCTGATCACCTCCGAC 672  
 DB 3127 AGCAAACTATAACATTTGAAATAAATAAATTCAGATACCGGTTTGATTTACTTCAGAT 3186  
 QY 673 TCCGACACATCAACATGTGTGATCCGTTGACTTCTACATCTTCCCAAGGAGTTGGAGCT 732  
 DB 3187 TCTGATACATCAATATGTGATAAGAGATTTTATATATTTTCTAAAGAAATAGATGCT 3246  
 QY 733 AAGACATCAACATCCCTGTTCAACTCTTCAGTACACCAACGCTCGTCAAGGACTACTG 792  
 DB 3247 AAGATATTAATATATTTAATAGCTTGCATATCTACTTGTGTAAGATTTATGG 3306  
 QY 793 GGTACAGGCTGAGATACACAAGGAGTACTACATGTGTCAACATCGACTACTTTGAACAGA 852  
 DB 3307 GGAATGATTTAAGATATATAAAGAAATATATATGTTTAAATAGATTTTAAATAGA 3366  
 QY 853 TACATGTAGCCCACTCCAGACAGATCGTTTCAACACCAAGAGTACACACGACTTC 912  
 DB 3367 TATATGTATGCAACTCACGACAAATTTTAAATACAGTGAATAATATAATGACTTC 3426  
 QY 913 AACGAGGTTTACAAGATCATCATCAAGGATCAGAGGTAAACCAACGACACCCAGAGTC 972  
 DB 3427 AATGAGGATATAAATTAATAAAGAAATCAGAGAAATACAAATGATAGTAGTA 3486  
 QY 973 AGAGTGTGACATCCTGTACTTCGACATGACTATCAACACAGGCTTACACCTGTTTC 1032  
 DB 3487 CGAGGAGGAGATATTTTATATTTTATGATGACAATTAATAACAAAGCATATAATTTGTTT 3546  
 QY 1033 ATGAAGAAGGAGACCATGTACGCGACAAACCTCCACCGAGGACATCTACGCCATCGGT 1092  
 DB 3547 ATGAAGATGAACATGTATGCGAGATTAATCATAGTACTGAAGATATATATGATAGGT 3606  
 QY 1093 CTGCGTGAGCAGACCAAGGATCAACGACAAACATCATCTTCGAGTCCGAGCCAGTGAAC 1152  
 DB 3607 TTAAGAGAACAAACAAAGGATATAAATGATATATATTTTCAATACAAACCAATGAAT 3666  
 QY 1153 RACACTTACTACTCGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAGAACATCTCC 1212  
 DB 3667 AATCTTATTTATGCAATCTCAATATTTAAATCAAAATTTAATGGAGAAATATTTCT 3726  
 QY 1213 GGTATCTGTCCATCGTACCTACAGATTCGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1272  
 DB 3727 GGAATATGTTCAATAGTACTATCGTTTGTAGCTTGGAGGTGATGGTATAGACAAAT 3786  
 QY 1273 TACTGGTTCCAACTGTCAGCAGGTAACCTACGCTTCTGCTGGAGTCCACTTCCACC 1332  
 DB 3787 TATTTGGTCCCTACTGTGAAGCAAGAAATTTATGCTTCATTTATAGATCAACATCAACT 3846  
 QY 1333 CACTGGGATTCGTCCTCCAGTCTCCGAGTAA 1362  
 DB 3847 CATTTGGGTTTGTACCTGTAAGTGAATAA 3876

RESULT 6  
 US-10-205-516-5  
 ; Sequence 5, Application US/10205516  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zhong, Jun  
 ; TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
 ; TITLE OF INVENTION: Neurotoxins Through Recombinant DNA Technique  
 ; FILE REFERENCE: JZBtX1  
 ; CURRENT APPLICATION NUMBER: US/10/205,516  
 ; CURRENT FILING DATE: 2002-07-25  
 ; NUMBER OF SEQ ID NOS: 28  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 3876  
 ; TYPE: DNA  
 ; ORGANISM: Clostridium botulinum  
 ; FEATURE:  
 ; NAME/KEY: CDS

LOCATION: (1)...(3876)  
PUBLICATION INFORMATION:  
AUTHORS: Hauser, D. F.  
TITLE: Organization of the botulinum neurotoxin C1 gene and  
TITLE: its associated non-toxic protein genes in Clostridium  
TITLE: botulinum C 468  
JOURNAL: Mol. Gen. Genet.  
VOLUME: 243  
ISSUE: 6  
PAGES: 631-640  
DATABASE ACCESSION NUMBER: GeneBank / X2793  
DATABASE ENTRY DATE: 1993-05-03  
RELEVANT RESIDUES: 1 TO 3876  
US-10-205-516-5

Query Match 48.5%; Score 665.2; DB 42; Length 3876;  
Best Local Similarity 68.3%; Fred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATCAACATCTCTCCACCAACAACTCCCTGTTGAAGGACATCATCAAC 72  
Db 2527 ACAATACCCCTTAATATTTTTCATATACTAATAATCTTTATTAAGAGATATTAAT 2586  
QY 73 GAGTACTTCACACATCAACAGATCCAGATCCCTGTCCTCCAGAACCGTAAGAACACC 132  
Db 2587 GAATATTTCAATAATATTAATGATTCAAAATTTTGAGCCCTACAAAACAGAAAATACT 2646  
QY 133 TTGGTGCACACCTCCCGTTACACGCGGAGGTCCTCCGAGGAGGTTGAGCTCCAGCTGAAC 192  
Db 2647 TTAGTGATACATCAGGATATATGACAGAGTGAGTGAAGAGGCGATGTTTCAGCTTAAT 2706  
QY 193 CCAATCTCCATTCACATCAAGCTGGGTTCCCTCGGTGAGGACAGAGGTAAGGTCAIC 252  
Db 2707 CCAATTTTCCATTTGACTTTAAATTAGTAGTTGACGAGGAGGATAGAGGTAAGTTATA 2766  
QY 253 GTACCCAGAACAGAACATCGCTACAACTCCATGCTAGAGTCCCTCTCCATCTCCCTC 312  
Db 2767 GTACCCAGAAATGAAATATGATATATATCTATGATGAAGTTTAGCATTAGTTT 2826  
QY 313 TGGATCAGAAATCAAAAGTGGTCTCCAACTTCCAGGTTACACCATCATCGCTCCGTC 372  
Db 2827 TGGATTAGAATAAATAAGTGGTAAGTAATTTACCTGGATATATTAATGATAGTGT 2886  
QY 373 AAGAACAACTCCGGTGGTCCATCGGTATCATCTCCAACTTCTCGTCTCCACCTGAAG 432  
Db 2887 AAAAAATCACTGGTGGATATAGTATATTTAGTAATTTTGTATTTTACITTAATA 2946  
QY 433 CAGAACGAGACTCCGAGAGTCCATCACTCTCCTAGCATCTCCAAACACGCTCCT 492  
Db 2947 CAAATGAAGATAGTGAACAAAGTAAATTTTAGTTATGATATATCAATTAATGCTCT 3006  
QY 493 GGTACAAAGTGGTCTTCGTCACCGTCACCAACACATGATGGTAAACATGAAGATC 552  
Db 3007 GGATACAAATAGTGGTGTGTAAGTGTATACATATGATGGGAATATGAAGAT 3066  
QY 553 TACATCAACGGTAAAGTGTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAACT 612  
Db 3067 TATATAAATGAAAATTAATAGATATATAAAGTTAAAGAACTAACTGGAATTAATTT 3126  
QY 613 TCCAAGACCATCACTCCAGATCAACAGATCCAGACACCGGTCTGATCACTCCGAC 672  
Db 3127 AGCAAACTATAACATTTGAAATAAATAAATTTCCAGATACCGGTTTGTACTTCAAT 3186  
QY 673 TCGGACAACATCAACATGGTGGATCCGTGACTTCTACATCTTCCCAAGGAGTTGACGGT 732  
Db 3187 TCTGATAACATCAATATGTTGATGAAGAGATTTTATATTTTGTAAAGAAATTAGATGGT 3246  
QY 733 AAGGACATCAACATCTGTTTCACTCTGCTGAGTACACCAACGCTCGTCAAGGACTACTGG 792  
Db 3247 AAGAGATTAATATATTTTAAATAGCTTGCATATACATAATGTTGTAAGAGATTATGG 3306  
QY 793 GGTAACGACCTGAGATACAAACAGAGGAGTACTACATGGTCAACATCGACTACTTGAACAGA 852

Db 3307 GGNAATGATTTAAGATATAATAAGAAATATATATATGTTTAAATAGATTATTTAAATAGA 3366  
QY 853 TACATGTACGCCCAACTCCAGACAGATCGTCTTCAACACCCAGCTAACAACAGCACTTC 912  
Db 3367 TATATGTATCGCAACTCAGCAAAATTTGTTTAAATACACGTAGAAATAAATAGCACTTC 3426  
QY 913 AAGCGGGTTACAAGATCATCATCAAGCGTATCAGAGGTAAACCAACGACACCAAGAGTC 972  
Db 3427 AATGAAGGATATAAATATATAAAGAAATCAGAGAAATACAATGTACTAGAGTA 3486  
QY 973 AGAGTGGTGACATCCCTGCTACTTCGACATGACTATCAACAACAGGCTTACACCTGTTTC 1032  
Db 3487 CGAGGAGGAGATATTTTATATTTTGTATGACATTAATAACAAGACATATAATTTGTTT 3546  
QY 1033 ATGAAGAACGAGACCATGTACGCCGACACCACTCCACCGAGGACATCTACGCCATCGGT 1092  
Db 3547 ATGAAGAATGAACATGATGATCGACATATCATATGATGATGATATATATGCTATAGGT 3606  
QY 1093 CTCGTGAGCAGCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCCAATGAAC 1152  
Db 3607 TTAAGAGAAACAACAAAGGATATAAATGATATATATTTTCAATACAAACCAATGAAT 3666  
QY 1153 AACACATCTACTACTACGCTTCCAGATCTTCAAGTCCAACTTCAACGGGTGAGAACATCTCC 1212  
Db 3667 AATCTATATTTACGCATCTCAATATTTAATCAATTTTATGAGAAATATTTCT 3726  
QY 1213 GSPATCTGTCCATCGGTACCTACGATTCCTGCTGGGTGGTGTGCTGTGATGATGATGATGAT 1272  
Db 3727 GGAATATGTTCAATAGTACTTATGCTTTAGACTTGGAGGTGATTTGTTATGATGATGAT 3786  
QY 1273 TACTTGTCTCCAACTGTCAAGCAGGAGTAACTACGCTCTTCTGCTGGAGTCCACTTCCACC 1332  
Db 3787 TATTTGTGCTTACGTGAGCAGGAAATTTAGCTTATTATTAGAAATCAATCAATCACT 3846  
QY 1333 CACTGGGATTCGTCCCACTCTCCGAGTAA 1362  
Db 3847 CATTGGGTTTGTACTCTGTAAGTGAATAA 3876

## RESULT 7

US-10-271-012-59  
Sequence 59, Application US/10271012  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
Thalley, Bruce S.  
TITLE OF INVENTION: Multivalent Vaccine For Clostridium  
Botulinum Neurotoxin  
NUMBER OF SEQUENCES: 82  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/271,012  
FILING DATE: 15-Oct-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/704,159  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPHD-02304  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3876 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3873  
; SEQUENCE DESCRIPTION: SEQ ID NO: 59:  
US-10-271-012-59

Query Match 48.58; Score 665.2; DB 43; Length 3876;  
Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;  
QY 13 ACCATCCCATTCACATCTTCTCTACCAACAACCTCCCTGTTGAGGACATCATCAAC 72  
DB 2527 ACAATACCCITTAATATTTTTCATATCTAATAATCTTTTATAAAGATATAATAT 2586  
QY 73 GAGTACTTCAACACATCAACGATCCAGATCTGTCCTCGAGAACCGTAAACACAC 132  
DB 2587 GAATATTTCAATAATTAATGATTCAAAAATTTTGAGCCTACAAACAGAAAAATCT 2646  
QY 133 TTGGTCGACACCTCCGGTTACAACGCCGAGGTCTCCGAGGAGGTGAGCTCAAGTGAAC 192  
DB 2647 TTAGTGGATACATCAGGATATATGAGAGAGTGTGAGAGAGCGGATGTCACCTTAAT 2706  
QY 193 COAATCTTCCATTCGACTTCAAGCTGGTTCCTCCGGTGAGGACAGAGGTAAAGTCATC 252  
DB 2707 CCAATATTTCCATTTGACTTTAAATAGTAGTGTGAGGAGGAGGAGGTAAGTTATA 2766  
QY 253 GTCACCCAGAGAGAGATCGTCTACAACCTCCAGTACGAGTCTCTCCATCTCTCTC 312  
DB 2767 GTAACCCAGATGAATATTTGTATATTAATCTATGATGAAGTTTATAGATTTT 2826  
QY 313 TGGATCAGAATCAACAGTGGGTCTCCAACTGCCAGTTTACACCATCAPCGACTCCGTC 372  
DB 2827 TGGATTAGATAAATAATAGGTAGTAATTTTACCTGGATATCTAATTTAGTAGTT 2886  
QY 373 AGAACAACCTCCGTTGCTCATCGGTATCATCTCCAACTTCTGGTCTCACCTGAAG 432  
DB 2887 AAAATAACTCAGGTGGAGTATAGGTATATATAGTAATTTTATAGTATTTACTTTAA 2946  
QY 433 CAGAACGAGGACTCCGAGCAGTCCATCACTTCCTTACGACATCTCCAAACAGCTCCT 492  
DB 2947 CAAATGAAGATAGTGAACAAAGTATAAATTTTATGATATATCAAAATATGCTCT 3006  
QY 493 GGTACACACAGTGGTCTTCTGTCACCGTCCACCAACACATGATGGTACATGAATC 552  
DB 3007 GGATCAATAAATGTTTGTGTAACCTGTTACTAACATATGATGGGAATATGAAGAT 3066  
QY 553 TACATCAACGTTAAGTCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCACTTC 612  
DB 3067 TATATAATGGAATAATATAGATCTATAAAGTTTAAAGAACTAACTGGAATTAATTT 3126  
QY 613 TCCAGACATCACTCCGATCGATCAACAGATCCGACACCGGTCTGATCACCCTCGGAC 672  
DB 3127 AGCAAACTATAACATTTGAAATAAATAAATCCAGATACCGGTTTGTATCTTCAGAT 3186  
QY 673 TCCGACACATCAACATTTGGATCGGTGACTTCTACATCTTCGCAAGGAGTTGGACGGT 732  
DB 3187 TCTGATACATCAATATGGAATGAAGATTTTATATATTTGCTAAGATATAGATGGT 3246  
QY 733 AAGGACATCAACATCTCTTCACTCTTGCAGTACACCAACGCTCGTCAAGGACTACTGG 792  
DB 3247 AAAGATATAATATATTTTAAATAGTGTGAATATCTAATGTTGTTAAAGATTTTGG 3306  
QY 793 GGTACGACCTGAGATACAAACAGGAGTACATGCTGCAACATCGGACTACTTGAACAGA 852  
DB 3307 GGAATGATTTAAGATATAAATGAAGATATATATGTTTATATAGATTTTAAATAGA 3366

QY 853 TACATGTACGCCAACTCCACAGACATCGCTTTCACACACAGAGCTTAACAAACAGACTTC 912  
DB 3367 TATATGTATGCGAACTCAGCAAAATTTGTTTTTAATACACGCTAGAAATTAATATGACTTC 3426  
QY 913 AACGAGGTTACAGATCATCATCAAGCGGTATCAGAGGTAAACACCAACGACACAGAGTC 972  
DB 3427 AATGAAGATATAAATTAATAAAGAGATCAGAGGAATACAAATGATACTAGAGTA 3486  
QY 973 AGAGTGGTGACATCTCTGTTACTTCGACATGACTATCAACAAAGGCTTACAACTGTTTC 1032  
DB 3487 CGAGGAGGAGATATTTTATATTTTGTATGACAAATTAATAACAAAGCATATAATTTGTTT 3546  
QY 1033 ATGAAGAACGAGACCATGTAGCGCGCAACCACTCCACCGAGGACATCTACGCCATCGGT 1092  
DB 3547 ATGAAGATGAATATGTATGAGATATATCATATGATGATGATATATATGCTATAGT 3606  
QY 1093 CTGGTGAGCAGACCAAGGACATCAACGACACATCATCTTCCAGATCCAGGCAATGAAC 1152  
DB 3607 TTAAGAGAACAAACAAAGGATATAAATGATATATATATTTTCAAAATACAAAGCAATGAAT 3666  
QY 1153 AACACTTACTACTACGCTCCGAGATCTTCAAGTCCAACTTCAACGCTGAGACATCTCC 1212  
DB 3667 AATCTTATTTATGAGCATCTCAAAATTTTAAATCAAAATTTTATGGAGAAAATATTTCT 3726  
QY 1213 GGTATCTGTTCCATCGGTACTTACAGATTCGCTCTGGTGGTGGTACTGTTACAGACACAAC 1272  
DB 3727 GGAATATGTTCAATAGTACTTATCGTTTTAGACTTTGGAGGTGATTTGATAGACAAAT 3786  
QY 1273 TACTTGGTCCAACTGTCAGACAGGTTAACTAGCCCTCTTCTGCTGGAGTCCACTTCCACC 1332  
DB 3787 TATTGTGCTACTGTGAAACAGGAATTTATGCTTCATTTATAGATCAACATCAACT 3846  
QY 1333 CACTGGGAGTTCGCCAGTCTCCGAGTAA 1362  
DB 3847 CATTTGGGTTTGTACCTGTAAAGTGAATAA 3876

## RESULT 8

US-08-954-302-2  
; Sequence 2, Application US/08954302  
; GENERAL INFORMATION:  
; APPLICANT: Lance Simpson, Nikita Kiyatkin,  
; APPLICANT: Andrew Maksymowich  
; TITLE OF INVENTION: Compositions and Methods for Systemic  
; TITLE OF INVENTION: Delivery of Oral Vaccines and Therapeutic Agents  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jane Massey Licata, Esq.  
; STREET: 66 E. Main Street  
; CITY: Marlton  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM 486  
; OPERATING SYSTEM: WINDOWS FOR WORKGROUPS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/954,302  
; FILING DATE: herewith  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane Massey Licata  
; REGISTRATION NUMBER: 32,257  
; REFERENCE/DOCKET NUMBER: JEFF-0164  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (609) 779-2400  
; TELEFAX: (609) 810-1454

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3950  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
ANTI-SENSE: NO  
US-08-954-302-2

Query Match 48.5%; Score 665.2; DB 13; Length 3950;  
Best Local Similarity 68.3%; Pred. No. 4.7e-130;  
Matches 922; Conservative 0; Mismatches 428; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACACATCTCCCTACACCAACAACTCCCTGTTGAAGGACATCATCAAC 72  
Db 2558 ACAATACCCCTTAAATATTTTTCATATACTAATAATCTTTTATTAAGAGATATAATTAAT 2617  
QY 73 GAGTACTTCAACACATCAACAGCTCCAGATCCCTGTCGCAACCGTAAGAACACC 132  
Db 2618 GAATATTTCAATATATTAATGATCAAAAATTTGAGCCTACAAAACAGAAAATACT 2677  
QY 133 TTGGTCGACACCTCCCGTTACAAACCGGAGTCTCCGAGGAGGTGACGTCCAGCTGAAC 192  
Db 2678 TTAGTGGATACATCAGATATATATGCAAGTGAAGTGAAGAGGCGATGTTTCAGCTTAAT 2737  
QY 193 CCAATCTCCATTCACACTCAAGCTGGGTTCTCCGGTGAGGACAGAGTAAGTCAATC 252  
Db 2738 CCAATATTTCCATTTGACCTTAAATAGTAGTCTCAGGGAGGATAGAGTAAAGTTATA 2797  
QY 253 GTCAACCCAGAACGAGAACATCGCTACAACTCCAGTACGAGTCCCTTCGATCCCTTC 312  
Db 2798 GTAACCCAGAAATATGTATATAATCTATGATGAAGTTTTCAGCATTAGTTT 2857  
QY 313 TGGATCAGATCAACAAAGTGGTCCCACTTCCAGGTACACCATCATCGACTCCGTC 372  
Db 2858 TGGATGAGATAAATAAATGGGTAAGTAATTTACCTGGATATACATAATGATAGTGT 2917  
QY 373 AAGAACAACCTCCGGTTGGCCATCGGTATCATCTCCAACTTCCCTGGTCTCCACCGTGAAG 432  
Db 2918 AAAAATAACCTCAGGTTGGATATAGTATATTTAGTAATTTTTCAGTATTTTAAAT 2977  
QY 433 CAGAACGAGACTCCGAGAGTCCACTCACTTCTCCTAGGACATCTCCAAACACGCTCCT 492  
Db 2978 CAAATGAAGATAGTGAACAAAGTAAATTTTAGTATGATATATCAATAATGCTCCT 3037  
QY 493 GGTATCAACAAAGTGGTTCTTCGTCACCGTCACCAACATGATGGTAAACATGAAGATC 552  
Db 3038 GGATACAATAATGGTTTTCGTAAGTGTACTAACAATATGATGGAAATATGAAGATT 3097  
QY 553 TACATCAACGGTAACTGATGACACCATCAGGTCAAGGATGACCGGTATCAACTTC 612  
Db 3098 TATATAATGGAAATTAATAGATCTATAAAGTTAAAGTAAGTAACTAACTGGAATTAATTT 3157  
QY 613 TCAAGACCATCACTCCGATGATCAACAGATCCCAAGACACCGGTCTGATCACTCCGAC 672  
Db 3158 AGCAAACTATACATTTGAATTAATAAATTCAGATACCGGTTGATTTACTTCAGAT 3217  
QY 673 TCGACAAACATCAACATGTTGATGCTGACTCTCACTCTCCGCAAGGAGTTGGACGGT 732  
Db 3218 TCTGATACATCAATATGTGGATAAGAGATTTTATATATTTTCCTAAAGAAATTAGATGGT 3277  
QY 733 AAGGACATCAACTCTGTTCACTCTTGCAGTACACCAACGCTGCTCAAGGACTACGG 792  
Db 3278 AAGATATTAATATATTTAATTAAGCTTGCAATATACATATGTTGTTAAGATTAATGG 3337  
QY 793 GGTAAACGCTGAGATACAAACAGGAGTACTACATGGTCAACATCGACTTCTTGAACAGA 852  
Db 3338 GGAAATGATTTAAGATATAAAGAATATTAATGTTTAATATAGATTATTTAAATAGA 3397  
QY 853 TAGATGTACGCCAACCCACAGATCGCTTCCAAACACGACGCTAACACACAGCACTTC 912  
Db 3398 TATATGTATCGGAACACTCACAGAAATGTTTTTAATACAGTAGAAATAATAATGACTTC 3457

QY 913 AACGAGGTTACAGATCATCATCAAGCTATCAGAGGTACACCAACGACACACAGATC 972  
Db 3458 AATGAAGGATATAAATATTAATAAAGAAATCAGAGGAAATACAAATGATCTAGAGTA 3517  
QY 973 AGAGGTGGTGACATCCTGTACTTCCAGATGACTATCAACACAGAGGCTACNACCTGTC 1032  
Db 3518 CGAGGAGGAGATATTTTATATTTGATGACAATTAATAACAAGCATATAATTTGTTT 3577  
QY 1033 ATGAAGAACGAGACCATGTACGCCGACCAACCACTCCACGAGGACATCTAGCCATCGGT 1092  
Db 3578 ATGAAGATGAACATATGTATGACATATATCATAGTACTAGAGATATATATGCTATAGGT 3637  
QY 1093 CTGCGTGACCAACCAAGGACATCAACGACACATCATCTTCCAGATCCAGCCAAATGAC 1152  
Db 3638 TTAAGAGAACAAACAAAGGATATAAATGATAATATATATTTCAATACAAACCAATGAAT 3697  
QY 1153 AACACTTACTACTAGCTTCCAGATCTTCAAGTCCAACTTCAACGGTGAGAACATCTCC 1212  
Db 3698 AATACTTATTAATACGATCTCAATATTTAAATCAAAATTTTAATGGAGAAATATTTCT 3757  
QY 1213 GGTATCTGTTCCATCGGTACCTACAGATTCGCTGGGTGGTGGTACTGGTACAGACAAAC 1272  
Db 3758 GGAATATGTTCAATAGGTACTTATCGTTTACACTTGGAGGTGATTGGTATAGACACAAT 3817  
QY 1273 TACTTGGTTCCAACTGTCAAGCAGGTAACTACGCTCCTTCTGGAGTCCACTTCCACC 1332  
Db 3818 TATTTGGTCCCTACTGTGAAGCAAGAAATTAATGCTTCATTTAATGAATCAACATCACT 3877  
QY 1333 CACTGGGATTCGTCCTCCAGTCTCCGAGTAA 1362  
Db 3878 CATTTGGGTTTTGTACCTGTAAGTGAATAA 3907

## RESULT 9

US-10-205-516-19  
Sequence 19, Application US/10205516  
GENERAL INFORMATION:

APPLICANT: Zhong, Jun  
TITLE OF INVENTION: A Method For Producing Biologically Active Botulinum  
TOXINS  
FILE REFERENCE: JZbtx1  
CURRENT APPLICATION NUMBER: US/10/205,516  
CURRENT FILING DATE: 2002-07-25  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 19  
LENGTH: 3906

TYPE: DNA  
ORGANISM: Clostridium botulinum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(3906)  
FEATURE:  
NAME/KEY: misc\_signal  
LOCATION: (1347)..(1358)  
OTHER INFORMATION: factor xa site  
FEATURE:  
NAME/KEY: misc\_signal  
LOCATION: (3886)..(3903)  
OTHER INFORMATION: 6-histidine tag

US-10-205-516-19

Query Match 48.4%; Score 663.2; DB 42; Length 3906;  
Best Local Similarity 68.4%; Pred. No. 1.2e-129;  
Matches 917; Conservative 0; Mismatches 423; Indels 0; Gaps 0;

QY 13 ACCATCCCATTCACATCTTCTCTACACCAACAACTCCCTGTTGAAGGACATCATCAAC 72  
Db 2539 ACAATACCCCTTAAATATTTTTCATATACTAATAATCTTTTATAAAGATATAATTAAT 2598  
QY 73 GAGTACTTCAACACATCAACGACTCCAGATCCCTGCTCCGACGAGTAAAGAACACC 132  
Db 2599 GAATATTCAATAATAATTAATGATTCAAAAATTTTGAGCCTACAAAACAGAAAAATACT 2658

Qy	1213	GGTATCTGGTTCCATCGGTACCTTACAGTTCGGTCTGGGGTGGTACTGGTACAGACACAAAC	1217
Db	3739	GGAAATATGTTTCAATAGTAGTACTTATCGTTTATAGACTTTGGAGGTGATTTGGTATATAGACAAAT	3798
Qy	1273	TACTTGGTTCCTCAACTCTCAAGCAGGTAACACTAGCGTCCCTCTCGTGGAGTCCACACTTCCACC	1332
Db	3799	TATTTGGTGCCTACTCTGAAGCAAGGAAATATATGCTTCATTTATTAGTAATCAACATCAACT	3858
Qy	1333	CACTGGGGATTCTGCCAGT	1352
Db	3859	CATTGGGGGTTTTGTACCTGT	3978
RESULT 10			
US-09-730-237-4			
; Sequence 4, Application US/09730237			
; GENERAL INFORMATION:			
; APPLICANT: Walker, Patricia			
; TITLE OF INVENTION: Methods of Treating Hyperhidrosis			
; FILE REFERENCE: 2933			
; CURRENT APPLICATION NUMBER: US/09/730,237			
; CURRENT FILING DATE: 2000-12-05			
; NUMBER OF SEQ ID NOS: 12			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 4			
; LENGTH: 3876			
; TYPE: DNA			
; ORGANISM: botulinum toxin			
US-09-730-237-4			
Query Match 48.3%; Score 662; DB 29; Length 3876;			
Best Local Similarity 68.1%; Pred. No. 2.2e-129;			
Matches 920; Conservative 0; Mismatches 430; Indels 0; Gaps			
Qy	13	ACGATCCATTCACATCTCTCTCTACACGACAACTCCCTGTTGAAGGACATCATCAAC	72
Db	2527	ACAAATCCCTTTAATATTTTTTCATATACTATAATCTTTATTAAGAGATATAATTAAT	2586
Qy	73	GAGTACTTCAACACATCAAGCACTCCAAAGATCCTGTCCTCGACGAAACCGTAAGAACACC	132
Db	2587	GAATATTTCAATAATAATTAATGATTCAAAAATTTTGAGCCTACAAAAACAGAAAAATACT	2646
Qy	133	TTGTTCCACACCTCCGGTTACACGCGGAGGTCTCCGAGGAGGTTGAGTCCAGCTGNAAC	192
Db	2647	TTAGTGATACATCAGGATATAATGACGAAGTGAGTGAAGAGGCGATGTTCCAGCTTAAT	2706
Qy	193	CCAACTTCCCATTCGACTTCAAGCTGGGTCTCTCGGTGAGGACAGAGTAAGTCAATC	252
Db	2707	CCAAATTTCCATTTTGAATTTAATAGGTAGTTCAGGGGAGGATAGAGGTAAGTTATA	2766
Qy	253	GTCACCCAGACAGACATCTGCTTACAACTCCATGTCAGAGTCCCTCTCTCCATCTCCTTC	312
Db	2767	GTAACCCAGAATGAANAATATGTATATAATCTATGTATGAAAGTTTTAGCATTAGTTTT	2826
Qy	313	TGGATCAGAATCAACAGTGGGTCTCCAACTTGCAGGTTACACCATCATCGACTCCGTC	372
Db	2827	TGGATTAGAATAAATAAATGGGTAAGTAATTTACOTGGATATACTATAATTGATAGTGT	2866
Qy	373	AAGAACAACTCCGGTTGGTCCATCGTATCACTCCAACTTCTTGCTTCCACCTGNAAG	432
Db	2887	AAAAAATCACTCAGGTGGAGTAGGTATATTTAGTAAATTTTTTTAGTATTTACTTTAAA	2946
Qy	433	CAGAACGAGGACTCCGAGCAGTCCATCAACTTCTCTAGCACATCTCCAAACAGCTCCT	492
Db	2947	CAAAATCAAGATAGTGAACAAAGATAAATTTTAGTTATGATATATCAAAATATGCTCCT	3006
Qy	493	GGTTACACAAAGTGGTCTTCGTCGCGGTCCACACACATCATGGTAAACATGAAGATC	552
Db	3007	GGATAGAATAAATGGTTTTTTTGTACTGTACTAACATATATGATGGGAATATGAAGAT	3056
Qy	553	TACATCAACGGTAAGCTGATCGACACCATCAAGGTTCAAGGAGTTGACCGGTATCAACTTC	612
Db	3067	TATATAATGGAAAAATTAATAGATACTATATAAGATTAAAGAACTAACTGGGAATTAATTT	3126

QY	613	TCCAAGACCATCACCTTCGAGATCAACAAAGATCCAGACACCGGTCTGATCACTCCGAC	672
Db	3127	AGCAAAACTATACCATTTGAAATAAATAAATTCAGATACCGTTGATTACTTCAGAT	3186
QY	673	TCCGACAAACATCAACATGNGGATCGGTGCACCTCTACATCTTCGCCAAGGAGTTGGAGGT	732
Db	3187	TCGATGATACATCATATGCGGATAGAGATTTTATATATTTGCTAAGAATTAGATGGT	3246
QY	733	AAGGACATCAACATCCTGTTCAACTCCTTGCAGTACCAACGTCGTCGAAGGACTACTGG	792
Db	3247	AAAGATATTAATATATTATTTAATAGCTTGCAAATATACTAATGTGTGAAAGATTTATGG	3306
QY	793	GSFAACGCTGAGATACAAACAGGAGTACTACATGCTACATCGACTACTTGAACAGA	852
Db	3307	GGAAATGATTTAAGATATTAATAAGAAATATTATATGTTTAAATAGATTTTAAATAGA	3366
QY	853	TACATGTACGCCAACTCCAGACAGATCGTCTTTCAACACGACAGCTACACAACAGACTTC	912
Db	3367	TATATGTATGCGAACTCACGACAAATTTGTTTTAATCACGTAGAAATATAATGACTTC	3426
QY	913	AACGAGGGTTACAGATCATCATCAAGCGTATCAGAGTATACCAACGACACACGAGTC	972
Db	3427	AATGAAGGATATAAAATTATATAAAAGAATCAGAGAAATACAAATGATCTAGAGTA	3486
QY	973	AGAGGTGGTCACATCGCTTACTTCGACATGACTATCAACAACAGGCGCTACAACTGTC	1032
Db	3487	CGAGGAGGAGATATTTTATATTTTGNATGACAATTATACAAGCATATATTTGTTT	3546
QY	1033	ATGAAGAAGAGACCATGTACGCCGACACCACTCCACCGAGGACATCTACGCCATCGGT	1092
Db	3547	ATGAAGAAATGAAACTATGTATCGAATATCATCTAGTACTGAAGATATATATGCTATAGT	3606
QY	1093	CTCGGTGAGCAGACCAAGGACATCAACGACACATCATCTCCAGATCCAGCCAAATGAC	1152
Db	3607	TTAAGAGAACAAACAAGGATATAATGATATATATTTCAAATACAAACCAATGAAT	3666
QY	1153	AACACTTACTACTACGCTTCCAGATCTTCAAGTCCCACTTCAACGGTGAAGACATCC	1212
Db	3667	AATACTTATTTATGSCATCTCAAAATATTTAAATCAAAATTTAATGGAGAAATATTTCT	3726
QY	1213	GGTATCTGTTCCATCGGTACCTTACAGATTCGGTCTGGGTGCTGACTGGTACGACACAAC	1272
Db	3727	GGAAATGTTTCAATAGGTACTTATCGTTTTAGACTTGGAGGTGATGGTATAGACAAAT	3786
QY	1273	TACTTGGTTCCAACTGTCAAGCAGGGAACATACGCTCCTTGGCTGGAGTCCACTTCCACC	1332
Db	3787	TATTTGGTGCCTACTGTGAAGCAAGAAATATGCTTCAATTTAGATCAACATCACT	3846
QY	1333	CATCGGGATTCGTCCCACTCTCCGAGTAA	1362
Db	3847	CATTGGGGTTTTGTACCTGTAAAGTAAATA	3876

RESULT 11

US-10-051-952-4

00 10 051 032 4 ; Sequence 4, Application US/10051952

GENERAL INFORMATION:

APPLICANT: Walker, Patricia

; TITLE OF INVENTION: Methods of Administering Botulinum Toxin

; FILE REFERENCE: 2933CIP

; CURRENT APPLICATION NUMBER: US/10/051,952

; CURRENT FILING DATE: 2002-01-17

; PRIOR APPLICATION NUMBER: 09/730,237

; PRIOR FILING DATE: 2000-  
 : NUMBER OF SEC ID NOS: 12

```

; NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn Ver 2.1

```

```

; SOFTWARE: E
; SEQ ID NO: 4

```

```

: SEQ ID NO 4
: LENGTH: 3876

```

```

; LENGTH: 38
; TYPE: DNA

```

; LIFE: DNA  
; ORGANISM: botulinum toxin

US-10-051-952-4











Result No.	Score	Query Match	Length	ID	Description
1	95	6.9	135259	5	US-10-240-435-1585
2	58.2	4.2	3028	6	US-09-548-938A-2
3	50.6	3.7	2835	6	Sequence 2, Appli
C 3	50.6	3.7	3561	6	Sequence 1515, Ap
C 4	50.6	3.7	6414	6	Sequence 1685, Ap
5	50.6	3.7	6414	6	Sequence 1685, Ap
6	50.4	3.7	15645	6	Sequence 1626, Ap
7	48.4	3.5	1452	6	Sequence 111, App
8	47.6	3.5	783	6	sequence 1, Appli
C 8	46.8	3.4	8249	6	Sequence 1670, Ap
9	46.8	3.4	2456	6	Sequence 138, App
10	46	3.4	18034	6	Sequence 83, Appl
11	45.8	3.3	1092	6	Sequence 75, Appl
12	45.8	3.3	1821	6	Sequence 5, Appli
13	45.6	3.3	1821	6	Sequence 82, Appl
14	44.4	3.2	3401	6	Sequence 123, Appl
15	44.4	3.2	3401	6	Sequence 411, App
16	43.2	3.2	8833	6	Sequence 1410, Ap
17	42.8	3.1	2175	6	Sequence 85, Appl
18	42.8	3.1	3878	6	Sequence 907, App
19	42.8	3.1	4174	6	Sequence 5, Appli
20	41.8	3.0	2334	6	Sequence 129, App
21	41.8	3.0	2334	6	Sequence 129, App
22	41.8	3.0	2334	6	Sequence 129, App
23	41.8	3.0	2334	6	Sequence 129, App
24	41.8	3.0	2334	6	Sequence 129, App
25	41.8	3.0	2334	6	Sequence 129, App
26	41.8	3.0	2334	6	Sequence 129, App

```
QY 489 TCCGGTTACAAAGTGGTTCTTCGTCCCGTCCACCAACAAAGTGGTAAACATGAA 548
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69646 CACCACGATCAACAA-----CGTCACCAACCAACCAACCACTGTGATCAGCACCAT 69696
QY 549 GATCTACATCAAGCGTAAGCTGATCGACACCATCAAGGTCAAGGAGTTGACCGGTATCAA 608
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69697 CACCACCATCTACTACACCATCAACCAACCATGATCAGCATATCACCACCAACAC 69756
QY 609 CTTCTCAAGACCATCACTTCGAGATCAACAAGATCCAGACACCGGTCCTGATCACCTC 668
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69757 CATCACTGTGATCAGCACCATCAACCAACCATATTAGCACCATCAACCAACCATCAACAA 69816
QY 669 CGACTCCGACAAATCAACATGTGGTCCGTGACTTCTACATCTTCGCCAAGAGTTGGA 728
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69817 CGTCACCAACCAACCATCACTGATGAGCACCATCAACCAACCATCAACCAACCATGAT 69876
QY 729 CGGTAAAGACATCAACATCCTGTTCACCTCTTGCAGTACACCAACGTGCTCAAGGACTA 788
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69877 CACCATTTACCAACCAACCAACCATCAAGGTGATGACCAACCATCAACCAACCAT 69936
QY 789 CTGG--GGTAAAGACCTGAGATACACAGGAGTACTACATGGTCAACATCGACTACTT 845
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69937 CATGATCACCATCTCACCACCAACCATCACTGATGATCAGCACCATCATCACCACCAT 69996
QY 846 GAACAGATACATGTACGCCAACTCCAGACAGATCGCTTCAACACACAGACGTAACAACAA 905
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 69997 CACCACCAACCATGATGCCATTATCACCATCACTGTGATCAGCACCATCAACCAACCAT 70056
QY 906 CGACTTCAACGAGGTTTACAGATCATCATCAAGCGTATCAAGGTAAACCAACGACAC 965
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70057 CACCATATCACCACCAATATCACCATTATCATCACCACCATCAACCATCACTGTGGTCTG 70116
QY 966 CAGAGTCAGAGGTTGATCACTCTGACTTCGACATGATATCAACAACGAGCGCTACAA 1025
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70117 CACCATCACCATCATCACCATCATATTACAACCATCAACCATCAACCATCATCACCAT 70176
QY 1026 CTTCTTATGAGACAGACCATGATGACCGGACACACCATCGACCGGAGGATCTAGCC 1085
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70177 CACCACCAACCATCACTGTGATGACACCAACCATATTACCAACCATCAACCATCAAC 70236
QY 1086 CAT 1088
      |||||
Db 70237 CAT 70239
```

```
RESULT 2
US-09-548-938A-2
; Sequence 2, Application US/09548938A
; GENERAL INFORMATION:
; APPLICANT: EMALFARB, MARK AARON
; APPLICANT: BURLINGAME, RICHARD PAUL
; APPLICANT: OLSON, PHILIP TERRY
; APPLICANT: SINITSYN, ARKADY PANTELEIMONOVICH
; APPLICANT: PARRICHE, MARTINE
; APPLICANT: BOUSSON, JEAN CHRISTOPHE
; APPLICANT: PYNNONEN, CHRISTINE MARIE
; APPLICANT: PUNT, PETER JAN
; APPLICANT: VAN-ZEIJL, CORNELIA MARIA JOHANNA
; TITLE OF INVENTION: TRANSFORMATION SYSTEM IN THE FIELD OF FILAMENTOUS FUNGI
; FILE REFERENCE: 3123-4001
; CURRENT APPLICATION NUMBER: US/09/548,938A
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 3028
; TYPE: DNA
; ORGANISM: Chrysosporium lucknowense
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (46)..(47)
; OTHER INFORMATION: a, t, c, g, other or unknown
; FEATURE:
```

```
; NAME/KEY: modified_base
; LOCATION: (374)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-09-548-938A-2

Query Match      4.2%; Score 58.2; DB 5; Length 3028;
Best Local Similarity 44.3%; Pred. No. 4.4e-06;
Matches 292; Conservative 0; Mismatches 358; Indels 9; Gaps 1;

QY 364 GACTCGCTCAAGAACAACTCCGGTTGGTCCATCGGTATCATCTCCAACTTCTTGGTCTTC 423
Db 1729 GCCAAGGGCAAGCTCATCGCGGCCACACCCCTCTCTGSCACTCTCAGTGGCGGAGTGG 1788
QY 424 ACCTGAGACAGAACAGAGACCCGAGAGTGCATCAACTTCTCTAGACATCTCCAAAC 483
Db 1789 GTGAGAAATCAATCAACGCGCAACACCTTGACCCAGGTATCGAGAACCACTCCACC 1848
QY 484 AAGCTCTCTGGTTACAAAGTGGTTCTTCGTACCGTCCACCAACAAACATGATGGGTAA 543
Db 1849 CTGTCTACTCGTTACAAGGGCAAGATCTCTCACTGGGAGCTGTTAAGAGATCTTTGCC 1908
QY 544 ATGAGATCTTACATCAACGGTAACTGATCGACACCATCAAGGTCAAGGAGTTGACCGGT 603
Db 1909 GAGGAGCGCTCGCTCCGCGACAGCGCTTTCAGCCGCGTCTCGGGGAGGACTTTGTGCG 1968
QY 604 ATCAACTTCTCCAAAGACCATCACTTTCGAGATCAACAAGATCCACAGACACCGGTCTGATC 663
Db 1969 ATCGCTTTCGCGCGCGCGCGCGCGCGGATCCCAAGCGCAAGCTCTACATCAACGACTAC 2028
QY 664 ACCTCGACATCCGACAAACATCAACATGTGGATCGGTGACTTCTATCTTTCGCCAAGGAG 723
Db 2029 AACTCGACATTCGCAACTACGCAAGGTGACCCGGGATGTGCGAAGAGTCAACAAAG 2088
QY 724 TTGACCGGTAAAGACATCAACATCTCTGTTCAACTCTTTCAGTACACCAACGTCGTCAAG 783
Db 2089 TGGATCGCCAGGGCATCCGATCGAGCGCATCGCACCCAGTGGCCACTGCGCGGGCCC 2148
QY 784 GACTACTPGGGGTA-----ACGACCTGAGATACAAAGAGTACTATGTTCAAC 834
Db 2149 GGGGGGTGAACACGCGCGCGCGCGCGCGCGCTCAAGGGCGCTCGCGCGGCCAAC 2208
QY 835 ATCGACTACTTGAACAGATACATGTACGGCAACTCCAGACAGATGCTTTCAAACACCAAGA 894
Db 2209 GTCAGGAGATCGCCATCACCAGCTCGACATCGCGGGCGCTCCGCCAAGCACTACCTC 2268
QY 895 CGTAACAAACAGACTTCAACGAGGTTTACAGATCATCATCAAGCGTATCAGAGGTAAAC 954
Db 2269 ACCGTATGAACGCTCGCTCCAGGTCTCCAAAGTGGTGGCATCACCCTCTGGGGCGTC 2328
QY 955 ACCAAGACACACAGATGTCAGAGGTGGTGACATCTCTTACTTCGACATGACTATCAACAA 1013
Db 2329 TCTGACAAGACAGCTGGAGGTGCGAGCAGCAACCCGCTCTCTTCGACAGCACTACCA 2387

RESULT 3
US-10-092-411A-1515/c
; Sequence 1515, Application US/10092411A
; GENERAL INFORMATION:
; APPLICANT: Lynn Dcurette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-101
; CURRENT APPLICATION NUMBER: US/10/092,411A
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5676
; SEQ ID NO 1515
; LENGTH: 2835
```

```

; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-092-411A-1515

```

Query Match	3.7%	Score 50.6;	DB 6;	Length 2835;
Best Local Similarity	41.8%	Pred. No. 0.00041;		
Matches 378;	Conservative	0;	Mismatches 524;	Indels 3;
Gaps 1;				

QY	286	ATGTACGAGTCTCTTCCATCTCTCTTGTGATCAAGATCAACAAGTGGGTCTCCAACTTG	345
Db	2654	ATCAGCGAGTAGTGTGACATCTGTAGTGCATCAGAGTCACGAAGAAGTAGTGTAGTCAAC	2595
QY	346	CCAGGTTACACATATCATCACTCCGTCAAGAACAACTCCGGTTGGTCCATCGGTATCATC	405
Db	2594	GAGTAAAGTAGCTCATTAAGTGAATCGACAAGCACAGCGTTTCAGATTCACGCAACGGT	2535
QY	406	TCCAACTTCTGGTCTTTCACCTGGAAGCAGAACGAGGACTCCGAGCAGTCCCATCAACTTC	465
Db	2534	GTCAAAGTTCAGAAAGTGCCTCAACATCAACAGTGTGAGTGGCTCAACAAGTACAAGCAT	2475
QY	466	TCCTACGACATCTCCAACAACGCTCTGTTTACAACAAGTGGTTCTTCGTCAACGCTCAAC	525
Db	2474	TTCAGACTCAACGAGCACATCAGCTCTGTATAGCCATCAATAAAGCGAGCGAATCAGC	2415
QY	526	AACAACATGATGGGTAAACATGAAGATCTACATCAACGGTAAGCTGATCGACACCATCAAG	585
Db	2414	AAGCAAGTAAACTATTATCAGAATCAGTGAAGTACGTCGACGCTGTATAGCGCATCAAC	2355
QY	586	GTCAAGGAGTTGACCGGTATCAACTTCTCCAGAACCATCACTTCGAGATCAACAAGATC	645
Db	2354	GTCAACGAGCGTAAAGTGAATCCAAATAGCGCAAGTACGTCNTTAAGTAAATCGCAAGCAC	2295
QY	646	CCAGACACCGGTCTGATCACTCCGACTCCGACAAACATCAACATGTGGATCCGTGACTTC	705
Db	2294	AAGCGTTTCAGATTCAACAAGTACATCGACATCCGACAGTGGCTCCACATCAACGAGTGA	2235
QY	706	TACATCTTCCCAAGAGGTGTGACGGTAGGACATCAACATCTCTGTTCAACTCTCTCGAG	765
Db	2234	GAGTAGTCAGACAGTGCAGTAGTACGATCAATTAAGTGAATCGACAAGTACAGCGTTTCAGA	2175
QY	766	TACACCACGCTCGTCAAGGACTACTGGGTAACGACCTGAGATCAACAAGGAGTACTAC	825
Db	2174	TTCAACGAGTAGCTGCACATCCGACATCCGACAGTGCATCAATGTACGAGCGAGAGTGAATCAA	2115
QY	826	ATGTCACATCACTACTTTGAACAGATACATGTACGCCAAGTCCAGACAGATCGTCTTC	885
Db	2114	CAGTAAAGTAGCTCATTAAGTGAATCGACAAGTACGAGTCTTTCAGGCTCAAGGAGTGC	2055
QY	886	AACACCAGAGCTTAACAACAACGACTTCAACGAGGTTTACAAGATCATCATCAACGCTATC	945
Db	2054	GTGACATCCGACAGTGCATCAACGTCAACGAGTCAGAGTGAGTCAGACAGTACAAGTAC	1995
QY	946	AGAGTAAACACCAACACACAGCAGTCAGAGTGTGACATCTCTGTACTTCGACATCACT	1005
Db	1994	GTCAATTAAGTGAATGACAAAGTACGAGTCTTTCAGGCTCAACGAGTGGCTGCAGCATCCGA	1935
QY	1006	ATCAACAACAAAGGCTTACAACCTGTTTCATGAAGAACGAGACCAATGTACCGCGACAACAC	1065
Db	1934	CAGTGCATCAACGTCAACAAGCGAGAGTGAATCAACAAGTGAAGTACATCTGTTGAGTGA	1875
QY	1066	TCCACCGAGGACATCTACGCCATTCGGTCTGCGTAGCAGACCAAGGAGCATCAACAGCAAC	1125
Db	1874	ATCATTAAGCACACA---AGGGTTTCAGATTCAACAAGCGCGTCAACGTCAGAAAGTGCATC	1818
QY	1126	ATCATCTTCCAGATCCAGCGCAATGAACAACACTTACTACTAGGTTTCCACAGATCTTCAAG	1185
Db	1817	CACATCAACGAGTGAGGTGAGTCAACACAGTCCGACACATCGTTAAGTGAATCATTTAAG	1758
QY	1186	TCCAA 1190	
Db	1757	TACAA 1753	

## RESULT 4

```

US-10-092-411A-1685/C
;
; GENERAL INFORMATION:
;
; APPLICANT: Lynn Doucette-Stamm et al
;
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO
;
; TITLE OF INVENTION: EPIDERMIDS FOR DIAGN
;
; FILE REFERENCE: 032796-101
;
; CURRENT APPLICATION NUMBER: US/10/092,411A
;
; CURRENT FILING DATE: 2002-03-07
;
; PRIOR APPLICATION NUMBER: US 09/134,001
;
; PRIOR FILING DATE: 1998-08-13
;
; PRIOR APPLICATION NUMBER: US 60/064,964
;
; PRIOR FILING DATE: 1997-11-08
;
; PRIOR APPLICATION NUMBER: US 60/055,779
;
; PRIOR FILING DATE: 1997-08-14
;
; NUMBER OF SEQ ID NOS: 5676
;
; SEQ ID NO 1685
;
; LENGTH: 3561
;
; TYPE: DNA
;
; ORGANISM: Staphylococcus epidermidis
;
US-10-092-411A-1685

```

Query Match	3.7%	Score 50.6	DB 6	Length 3561
Best Local Similarity	41.8%	Pred. No. 0.00046		
Matches	378	Conservative	0	Mismatches 524
			Indels	3
			Gaps	
Qy	286	ATGTPACGAGTCCTTCCACTCTCCTCTTCGATCAGAAATCAACAAGTGGGTCTCCAACTTG	345	
Db	3450	ATACGCGAGTACGTCGATCTCTGATAGTCATCAGAGTCACGCAAGAAGATGAGTCAAC	3391	
Qy	346	CCAGGTTTACACCATTCGACTCCGTCACAGACAACTCCGGTTGGTTCATCGGTATCATC	405	
Db	3390	GAGTAAAGTACGTCATTAACTGAATCGACAAGCACAAAGCGTTTCAGATTACGAAGCGT	3331	
Qy	406	TCCAACTTCCTGGTCTTCAACCTGGAAGCAGAGAGGAGCTCCGACAGCTCATCACTTC	465	
Db	3330	GTCAACATCGAAGAAGTGGCTCAACATCAACAAGTGTGAGTGGCTCAACAAGTACAAAT	3271	
Qy	466	TCCTAACGACATCTCCAAACAGCGTCTCTGGTTTACAACAAGTGGTTCCTTCGTCACGGTCA	525	
Db	3270	TTACAGCTCAACGAGCAGATCGAGCTCTGATAGCGCATCAATAAAGCGAGGCAATCAGC	3211	
Qy	526	AACAACTGATGGGTACATGAAAGATCTTACATCAACGGTAAGCTGATCGACACCATCAAG	585	
Db	3210	AAGCACAAGTAAACTATTATCAGAACTCAGTGAAGTACGTCGAGCTGTATAGCGCATCAAC	3151	
Qy	586	GTCAGAGAGTTGACCGGTATCACTTCTCCAGACCATCAACCTTCGAGATCAAGAAGATC	645	
Db	3150	GTCAACGAGCGTAAAGTACGTCCTCAATAGCCGAATAGCTCTTAAGTAAATCGACAGCAC	3091	
Qy	646	CCAGACACCGGTCTGATCACTCCGACTCCGACACATCAACATGTGGATCGGTGACTTC	705	
Db	3090	AAGGTTTCAGATTCAACAGTACATCGACATCCGACAGTGGCTCCATCAACAGATGA	3031	
Qy	706	TACATCTTCGCGAAGAGTTGGAGGTAGGACATCAACATCCTGTTCAACTCTTCGAG	765	
Db	3030	GAGTGAATCGACAGTGCAGTACGTCATTAAAGTGAATCGCAAGTACAGCGTTTCAGA	2971	
Qy	766	TACACCAACGTCGTCAAGGACTACTGGGGTAACGACCTGAGATACAAAGAAGTACTAC	825	
Db	2970	TTCAACGAGTACGTCGCATCCGACAGTGCATCATGTGACGGAGGAGTGAATCAA	2911	
Qy	826	ATGTCAACTCGACTACTTGAACAGATACATGTAGCGCCAACTCCGACAGATCGTCTTC	885	
Db	2910	CAGTAAAGTACGTCATTAAAGTGAATCGCAAGTACGAGTCTTTTTCAGGCTCAACGAGTGC	2851	
Qy	886	AACACGAGCTTACACACAGCACTTCAACGAGGTTTACAAGATCATCATCAAGCGTATC	945	
Db	2850	GTCGACATCCGACAGTGCATCAACGCTACAGGTGAGGTGAGTGCAGCAGTACAGTAC	2791	
Qy	946	AGAGGTAACACCAACGACACCGACAGTCAAGGTTGGTGAATCCTCTACTTCGCATGACT	1005	

Db 2790 GTCATTAAAGTGAATCGACAAAGTACGAGTCTTTTCAGGCTCAACGAGTGCCTCGACATCCGA 2731  
 QY 1006 ATCAACGAAGAGGCTTACAACTGTTTCATGAAGAAGAGAGACCATGTACCGCCGACAAACCC 1065  
 Db 2730 CAGTGCATCAGCTCAACAGCGAGAGTGAATCAACAGTGAAGTACATCGTTGATGA 2671  
 QY 1056 TCCACGAGAGACATCTACGCGCATCGTCTCGTGAAGAGAGAGAGATCAACAGCAAC 1125  
 Db 2670 ATCATTAAAGCACA---AGCGTTTCAGATTCAACAAAGCGGTCAACGTCAGAAGTGCATC 2614  
 QY 1126 ATCATCTCCAGATCCAGCAACTGAACAACTTACTACTACGCTTCCAGATCTTCAAG 1185  
 Db 2613 CACATCAACGAGTGAAGTGAATCAACAGTGCACACATCGTTAAGTGAATTAAG 2554  
 QY 1186 TCCAA 1190  
 Db 2553 TACAA 2549  
 RESULT 5  
 US-10-092-411A-1626  
 ; Sequence 1626, Application US/10092411A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: US/10/092,411A  
 ; CURRENT FILING DATE: 2002-03-07  
 ; PRIOR APPLICATION NUMBER: US 09/134,001  
 ; PRIOR FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5676  
 ; SEQ ID NO 1626  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus epidermidis  
 US-10-092-411A-1626  
 Query Match 3.7%; Score 50.6; DB 6; Length 6414;  
 Best Local Similarity 41.8%; Pred. No. 0.00062;  
 Matches 378; Conservative 0; Mismatches 524; Indels 3; Gaps 1;  
 QY 286 ATGACGAGTCCCTTCCATCTCTCTGATCAGATCAACAAAGTGGTCCCACTTG 345  
 Db 2595 ATCAGCGAGTACGTCGACATCTGATAGTGCATCAGATCAGCAAGAAAGAGTGAAGTCAAC 2654  
 QY 346 CCAGGTACACCATCATCGACTCGGTCAAGAAACTCCGGTTGGTCCATCGGTATCATC 405  
 Db 2655 GAGTAAAGTACGTCATTAAGTGAATGATGCAAGCAAGCGTTTCAGATTCAGCAAGCGT 2714  
 QY 406 TCCAACTCTCTGGTCTTCAACCTGAAGCAAGAGAGTCCGAGAGTCCCACTCACTTC 465  
 Db 2715 GTCAAGGTCAAGAAAGTGGTCAACATCAACAAAGTGTGAGTGGCTCAACAGTACAGAT 2774  
 QY 466 TCCAGGACATCTCCACAGAGTCTCTGGTTTACAAAGTGGTCTTGTGTCGCTCAC 525  
 Db 2775 TCCAGATCAACGAGCAGATCGAGTGTGATAGCGCATCAATAAAGCGGAGCAATCAGC 2834  
 QY 526 AACACATGATGGGTAACTGAAGATGTACATCAACGTTAAGTGTATCGACACCATCAAG 585  
 Db 2835 AAGCACAAGTAACTATATCAGATCAGTGAAGTGTGATGAGTGGTCAAC 2894  
 QY 586 GTCAGGAGTGGGATGATCACTCTTCCAGACCATCACTTCGAGATCAACAGATC 645  
 Db 2895 GTCAGGAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2954  
 QY 646 CCAGACACCGGTGTGATCACTCGGATCCGACAAACATCAACATGAGTGGATCGGTGACTTC 705  
 Db 2955 AAGCGTTTCAGATTCAACAAAGTATCATCGACATCGGACATCGGTCGACATCAACAGATGA 3014

QY 706 TACATCTTCCCAAGAGGTTGGACGGTAAAGGACATCAACATCTCTTCAACTCTCTTCCGAG 765  
 Db 3015 GAGTGAAGTCAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3074  
 QY 766 TACACCAAGCTCTCGAGGCTACTGGGCTAAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 825  
 Db 3075 TTAACAGAGTACGTCGACATCCGACAGTGCATCAATGTGAGGAGGAGGAGTGAATCAAA 3134  
 QY 826 ATGCTCAACATCGACTACTTGAACAGATACATGTAGCCCAACTCCAGACAGATCGTCTTC 885  
 Db 3135 CAGTAAAGTACGTCATTAAGTGAATCGCAAGTACGAGTCTTTCAGGCTCAACGAGTGC 3194  
 QY 886 AACACAGAGCTTACAAACAGAGCTTCAACAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 945  
 Db 3195 GTCGACATCCGACAGTGCATCAACGTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3254  
 QY 946 AGAGTAAACACCAACGACACAGAGTGCAGAGTGGTGAACATCTCTTACATCTCTTACATCTCTTACATCTCT 1005  
 Db 3255 GTCATTAAAGTGAATCGACAGTACGAGTCTTTCAGGCTCAACGAGTGGTGCATCCGA 3314  
 QY 1006 ATCAACAAAGGCTTACAACTGTTTCATGAAGAGGAGACCATGTACCGCCGACAAACCC 1065  
 Db 3315 CAGTGCATCAACGCTCAACAAAGGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3374  
 QY 1066 TCCACCGAGGACATCTACGCGCATCGGTCTGCGTGAAGCAAGGAGATCAACAGCAAC 1125  
 Db 3375 ATCATTAAAGCACA---AGCGTTTCAGATTCACAAAGCGGTCAACGTCAGAAAGTGCATC 3431  
 QY 1126 ATCATCTTCCAGATCCAGCAACTGAACAGATCAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1185  
 Db 3432 CACATCAACGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3491  
 QY 1186 TCCAA 1190  
 Db 3492 TACAA 3496

RESULT 6  
 US-10-085-198-111  
 ; Sequence 111, Application US/10085198  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Alsobrook et al  
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
 ; FILE REFERENCE: 21402-279  
 ; CURRENT APPLICATION NUMBER: US/10/085,198  
 ; CURRENT FILING DATE: 2002-02-25  
 ; PRIOR APPLICATION NUMBER: 60/271,646  
 ; PRIOR FILING DATE: 2001-02-26  
 ; PRIOR APPLICATION NUMBER: 60/276,401  
 ; PRIOR FILING DATE: 2001-03-16  
 ; PRIOR APPLICATION NUMBER: 60/311,981  
 ; PRIOR FILING DATE: 2001-08-13  
 ; PRIOR APPLICATION NUMBER: 60/312,858  
 ; PRIOR FILING DATE: 2001-08-16  
 ; PRIOR APPLICATION NUMBER: 60/271,840  
 ; PRIOR FILING DATE: 2001-02-27  
 ; PRIOR APPLICATION NUMBER: 60/277,324  
 ; PRIOR FILING DATE: 2001-03-20  
 ; PRIOR APPLICATION NUMBER: 60/286,096  
 ; PRIOR FILING DATE: 2001-04-21  
 ; PRIOR APPLICATION NUMBER: 60/299,695  
 ; PRIOR FILING DATE: 2001-06-20  
 ; PRIOR APPLICATION NUMBER: 60/315,614  
 ; PRIOR FILING DATE: 2001-08-29  
 ; PRIOR APPLICATION NUMBER: 60/272,405  
 ; PRIOR FILING DATE: 2001-02-28  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 653  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 111  
 ; LENGTH: 15645  
 ; TYPE: DNA

ORGANISM: Homo sapiens  
US-10-085-198-111

Query Match 3.7%; Score 50.4; DB 6; Length 15645;  
Best Local Similarity 47.0%; Pred. No. 0.0011;  
Matches 190; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

QY 579 CATCAAGGTCAAGAGTGTACCGGTATCAACTTCCTCCAGACCATCACTTCGGAGATCAA 638  
DB 9784 CAGCAGGGTCAGGGCCTTATGGACACAGGCTGGTCACAGCCGACGACGACGACAA 9843

QY 639 CAAGATCCAGACACAC--GCTGTATCACTCCGACTCCGACACATCAATGTGAT 695  
DB 9844 CAACAGCACCAAGCAAGGGTCCATGGCAGGGCTGTCCTTTCAGCAAGTCTGATG 9903

QY 696 CCGTGACTTACATCTTCGCCAAGGAGTGGACGGTAAAGGACATCAACATCCTCTTCAA 755  
DB 9904 TCACAGAGTGGGAGCCCAACTGAGCGCTCAGCCCATGGGCTCTTACAGCAGGTTTCA 9963

QY 756 CTCCTTGCAGTACCAACAGTCTCAAGGACTACTGGGTACGACCTGAGATACAA 815  
DB 9964 CAGCAGCAGGTGCAACAGCAAGCAAGCACTTCAGCAGCAGCAGCAGCAGCTACAA 10023

QY 816 GGAGTACTACATGGTCAACATCGACTTACATGACATACATGTACGCCAATCCAGACA 875  
DB 10024 CAGCAACAGCACTTCAGCAGCAACAGCTTCACAGCAGCAACAGCAGCAGCTTCAA 10083

QY 876 GATCGTCTTCAACACAGCAGGTAAACAAACGACTTCAAGAGGGTTTACAGATCATCAT 935  
DB 10084 CAACAGCAGCAGCAACAGCTTCAACAGCAGCAACAGCAGCTTCAACAGCAGCAACAA 10143

QY 936 CAAGCGTATCAGAGGTAAACAAACAGCAACAGCAGAGTACAGAGTG 979  
DB 10144 CAACAGCAGCTTCAACAGCAGCAGCAGCAACAGCAGCAGTGG 10187

RESULT 7  
US-10-270-223-1  
; Sequence 1, Application US/10270223  
; GENERAL INFORMATION:  
; APPLICANT: BioImage A/S  
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR COMPONENTS  
; TITLE OF INVENTION: IMPACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMATION FROM  
; FILE REFERENCE: 3759-0126P  
; CURRENT FILING DATE: 2002-10-11  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1452  
; TYPE: DNA  
; ORGANISM: Aequoria Victoria and Human  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1452)  
; OTHER INFORMATION:  
US-10-270-223-1

Query Match 3.5%; Score 48.4; DB 6; Length 1452;  
Best Local Similarity 45.2%; Pred. No. 0.0011;  
Matches 218; Conservative 0; Mismatches 261; Indels 3; Gaps 1;

QY 624 CAGCTTCGAGATCAACAGATCCAGACACCGTCTGATCACTCCGACTCCGACATCAT 683  
DB 177 CAGCCTCGTACACCCCTACCTACGCTGCGGTGCTTACGCCCTACCCCGACCAT 236

QY 684 CAACATGTGATCCGCTGCTTACATCTTCGCCAAGGAGTGGACGTTAAGGACATCAA 743  
DB 237 GAAGCAGCAGGATCTTCAAGTCCGCCATGCCAGGCTACGTCAGGAGCGCACCAT 296

QY 744 CATCTGTTCATCTCTGAGTACACCAACGTCGTCAAGGACTACTGGGTGAAGCACT 803

DB 297 CTTCCTTCAAGGACGACGGCAACTACAAGACCCGCCCGAGGTGAAGTTTCGAGGGCGACAC 356

QY 804 GAGATACAACAAGGAGTACTACATGGTCAACATCGACTACTTGAACAGATACATGTACGC 863

DB 357 CTGTGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGCAACATCCITGG 416

QY 864 CAACTCCAGCAGATCGTCTTCAACACACAGAGTAAACAAACAGACTTCAACAGAGGGTTA 923

DB 417 GCA--CAAGCTGGAGTACAACATACAACAGCCACACAGCTCTATATCATGCGCGACAGCA 473

QY 924 CAAGATCATCATCAAGCGTATCAGAGGTAACCAACACACAGAGTCAGAGTGTGTA 983

DB 474 GAAGAAGGCGATCAAGGTGAATTCAGATCCGCCACAACTCGAGGACGGCAGCGTGCA 533

QY 984 CATCTGTACTTCGACATGACTATCAACAACAAGCCCTACAACCTGTTCATGAAGAAGA 1043

DB 534 GCTCGCGGACCACTACAGCAGACAGACCCCATCGGCGAGCGGCCCGTCTGCTGCCGA 593

QY 1044 GACCATGTACGCGGACCAACCACTCCACCGAGGACATCTACGCCATCGTCTGCGTGACA 1103

DB 594 CAACCACTACCTGAGCAGCCCGCTGAGCAAGACCCCAACAGAGAGAGCGCGATCA 653

QY 1104 GA 1105

DB 654 CA 655

RESULT 8  
US-10-092-411A-1670/c  
; Sequence 1670, Application US/10092411A  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-101  
; CURRENT APPLICATION NUMBER: US/10/092,411A  
; CURRENT FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: US 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5676  
; SEQ ID NO 1670  
; LENGTH: 783  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-10-092-411A-1670

Query Match 3.5%; Score 47.6; DB 6; Length 783;  
Best Local Similarity 42.8%; Pred. No. 0.0013;  
Matches 239; Conservative 0; Mismatches 319; Indels 0; Gaps 0;

QY 521 TCACCAACACATGATGGTAAACATCAAGATCTACATCAAGGTAAGCTGATCGACACCA 580

DB 610 TCAGATTCAACAAAGTACATCGACATCCGACAGTGGTCAACGTCACCAAGTGAGATGAC 551

QY 581 TCAAGGTCAAGGAGTTGACCGGTATCAACTTCTCCAAGACCATCACTTCCGAGATCAACA 640

DB 550 TCAACAGTGAAGTACATCGTTAAGTGTGAGTCGACAAAGCAAGCGTTTTCAGATTCAACA 491

QY 641 AGATCCAGACACCGGTCTGTGATCACTCCGACTCCGACAAACATCAACATGTGGATCGGTG 700

DB 490 AGCGCTCAACGTCAGCAAGTGCATCAACATCAACGAGCGTGTGAGTGTCCAAATAGCGCA 431

QY 701 ACTTCTACATCTTCGCCAAGGAGTTGGAGCTAAGGACATCAACATCTCTTCAACTCT 760

DB 430 AGTACGCTCATTAAGTGGCTCAACAGTACAAAGCTTTCAGATCAACGAGTACGTCGACA 371

QY 761 TGAGTACACCAAGTCTGTCAGGACTACTGGGTAAAGACCTGAGATACACAGGAGT 820

DB 370 TCAGCAAGTGCATCAACATCAACGAGCAGGAGCGACTCCGATAGCGCAAGTACGCTGTTA 311





```
Query Match 3.4%; Score 46; DB 6; Length 2466;
Best Local Similarity 43.2%; Pred. No. 0.0061;
Matches 269; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

QY 470 ACACATCTCCACACAGCTCTCTGTTACACAAAGTGGTTCTTCGTACCCGTCACCAACA 529
DB 1517 ACAGCAAGCGCGACATCGCGGTCTGTACGACATGGTTCGCACTGAGGACGCGCCCAACC 1576
QY 530 ACATGATGGTGAACATGAAGATCTACATCAAGGTAAGCTGATCGACACCATCAAGGTCA 589
DB 1577 GCACCAAGTTGTACGAGTTTACACAGCAGCGGAGGATTCACAGCCGGTCAAGGTCT 1636
QY 590 AGAGTTGACCGGTATCACTTCTCCACAGCATCACTTCGATGATCAACAGATCCAG 649
DB 1637 GGGACACACAGCAGCCCGCTCAAGAGCTGGAATGGCCCTCCAGCAAGCTGACCGTGG 1696
QY 650 ACACCGGTCTGATCACTCCGATCTCCGACAAACATCAACATGATGGATCGGTACTTCTACA 709
DB 1697 GCGACTTCGAGCGGCGGCAAGCCGACATCGCGTCTGTACGACATACGCGGAGGAG 1756
QY 710 TCTTCGCCAAGGAGTTGGAGGTAAGGATCAACATCACTCTGTTCACCTCTTGCAGTACA 769
DB 1757 GCGACCGCAGCCGTACGGGCTGTGGACCTTCCACAGCAGCCGTCGCGCTTCAACCGCC 1816
QY 770 CCAAGCTCGTCAAGGACTACTGGGTAAAGCACTGAGATACACAAAGGAGTACTACATGG 829
DB 1817 CCAAGCTGGTGGGACAGCAAC--AACGACCCGGTCAAGAGCTGGAAGTGAAGCGCA 1873
QY 830 TCAACATCGACTACTTGAACAGATACATGATGACCCAACTCCACAGACAGATCGCTTCAACA 889
DB 1874 GCAAGCCACCGTCGGGACTTCAAGGGGAGCGCAAGCGGACATCGCGTCTCTACG 1933
QY 890 CCAGACGTAAACAACAGACTTCAACGAGGTTTACAGATCATCATCAAGCGTATCAGAG 949
DB 1934 ACATGGTTCGACGAGGAGCGGCAACCGCACCAAGCTGTTCACCTTCAACCGGACGG 1993
QY 950 GTACACCAACAGACACAGATCAGAGTGGTGACATCTCTGATTCGACATGACTATCA 1009
DB 1994 CGACCGGTTTCAACAGCCCGGTCAAGGTGTGGACAGCAACAGCAGCCCGTGAAGAGCT 2053
QY 1010 ACAACAGGCTTCAACCTCTTCAAGACAGACAGACCATGATGACGCGACACCATCCA 1069
DB 2054 GGAATCGGAGCGCGTCAAGGTAGTGGAGGCGACTTCAACGGCGAGCGCAAGCGGACA 2113
QY 1070 CCGAGGACATCTAGCCATCGG 1091
DB 2114 TCGGGGTGTTAGGACTACGG 2135

RESULT 11
US-10-267-255-75
; Sequence 75, Application US/10267255
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 75
```

```
; LENGTH: 18034
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (302)...(302)
; OTHER INFORMATION: n is a or t or g or c
US-10-267-255-75

Query Match 3.4%; Score 46; DB 6; Length 18034;
Best Local Similarity 43.2%; Pred. No. 0.017;
Matches 269; Conservative 0; Mismatches 350; Indels 3; Gaps 1;

QY 470 ACACATCTCCACACAGCTCTCTGTTACACAAAGTGGTTCTTCGTACCCGTCACCAACA 529
DB 13031 ACAGCAAGCGCGACATCGCGGTCTGTACGACATGGTTCGCACTGAGGACGCGCCCAACC 13090
QY 530 ACATGATGGTGAACATGAAGATCTACATCAAGGTAAGCTGATCGACACCATCAAGGTCA 589
DB 13091 GCACCAAGTTGTACGAGTTTACACAGCAGCGGAGGATTCACAGCCGGTCAAGGTCT 13150
QY 590 AGAGTTGACCGGTATCACTTCTCCAGACCATCACTTCGATGATCAACAGATCCAG 649
DB 13151 GGGACACACAGCAGCCCGCTCAAGAGCTGGAATGGCCCTCCAGCAAGCTGACCGTGG 13210
QY 650 ACACCGGTCTGATCACTCCGATCTCCGACAAACATCAACATGATGGATCGGTACTTCTACA 709
DB 13211 GCGACTTCGAGCGGAGGAGCGGACATCGCGTCTGTACGACATACGCGGAGGAGG 13270
QY 710 TCTTCGCCAAGGAGTTGGAGGTAAGGATCAACATCACTCTGTTCACCTCTTGCAGTACA 769
DB 13271 GCGACCGCAGCCGTACGGGCTGTGGACCTTCCACAGCAGCCGTCGCGCTTCAACCGCC 13330
QY 770 CCAAGCTCGTCAAGGACTACTGGGTAAAGCACTGAGATACACAAAGGAGTACTACATGG 829
DB 13331 CCAAGCTGGTGGGACAGCAAC--AACGACCCGGTCAAGAGCTGGAAGTGAAGCGCA 13387
QY 830 TCAACATCGACTACTTGAACAGATACATGATGACCCAACTCCACAGACAGATCGCTTCAACA 889
DB 13388 GCAAGCCACCGTCGGGACTTCAAGGGGAGCGGCAAGCGGACATCGCGTCTCTACG 13447
QY 890 CCAGACGTAAACAACAGACTTCAACGAGGTTTACAGATCATCATCAAGCGTATCAGAG 949
DB 13448 ACATGGTTCGACGAGGAGCGGCGCAACCGCACCAAGCTGTTCACCTTCAACCGGACGG 13507
QY 950 GTACACCAACAGACACAGATCAGAGTGGTGACATCTCTGATTCGACATGACTATCA 1009
DB 13508 CGACCGGTTTCAACAGCCCGTCAAGGTGTGGACAGCAACAGCAGCCCGTGAAGAGCT 13567
QY 1010 ACAACAGGCTTCAACCTCTTCAAGACAGACAGACCATGATACGCGACACCATCCA 1069
DB 13568 GGAATCGGAGCGCGTCAAGGTAGTGGAGGCGACTTCAACGGCGAGCGCAAGCGGACA 13627
QY 1070 CCGAGGACATCTAGCCATCGG 1091
DB 13628 TCGGGGTGTTAGGACTACGG 13649

RESULT 12
US-10-270-223-5
; Sequence 5, Application US/10270223
; GENERAL INFORMATION:
; APPLICANT: Bioimage A/S
; TITLE OF INVENTION: AN IMPROVED METHOD TO DETECT INTERACTIONS BETWEEN CELLULAR CON
; TITLE OF INVENTION: INTACT LIVING CELLS, AND TO EXTRACT QUANTITATIVE INFORMA-TIO
; FILE REFERENCE: 3759-0126P
; CURRENT APPLICATION NUMBER: US/10/270,223
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1092
```

```
; TYPE: DNA
; ORGANISM: Aequoria Victoria and Human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1092)
; OTHER INFORMATION:
US-10-270-223-5

Query Match      3.3%; Score 45.8; DB 6; Length 1092;
Best Local Similarity 47.4%; Pred. No. 0.0045;
Matches 137; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 678 CAACATCAACATGTTGGATCCGTGACTTCTACATCTTCCGCAAGAGGTTGGACGGTAAGA 737
Db 291 CACCATCTTCTTCAAGGACGACGCGCACTACAAGCCGCGCGAGGTGAAGTTCGAGGG 350
QY 738 CATCAACATCTGTTCAACTCTTTCAGTACACCAAGCTGCTCAAGGACTACTTGGGTTAA 797
Db 351 CGACACCCCTGTTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGCGCAACAT 410
QY 798 CGACCTGAGATACACAGGAGTACTACATGGTCAACATCGACTACTTGNACAGATACAT 857
Db 411 COTGGGGCACAAGTGGAGTACAACACTACAACAGCCCAACAGCTGTATATCATGCGCGACAA 470
QY 858 GTACCCCAACTCCAGACAGATCGTCTTCAACACCAGACGTTAAACAACAACGACTTCAACGA 917
Db 471 GCAGAAGAAGCGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACGCGAGCGT 530
QY 918 GGGTTACAAGATCATCATCAAGCTGTTTCAGAGGTAAACCAACCAAGACAC 966
Db 531 GCAGCTCCGACCACTACCAGCAGACACACCCCATCGGCGAGCGGCC 579

RESULT 13
US-10-267-255-82
; Sequence 82, Application US/10267255
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mac, Y
; APPLICANT: Varcglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456Usl
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 82
; LENGTH: 1821
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-10-267-255-82

Query Match      3.3%; Score 45.6; DB 6; Length 1821;
Best Local Similarity 41.9%; Pred. No. 0.0066;
Matches 273; Conservative 0; Mismatches 379; Indels 0; Gaps 0;

QY 440 AGGATCCGACGATCCATCACTTCTCTTACGACATCTTCAACAAGCGTCTCGGTACA 499
Db 1031 AGCCAGCAAGCTCACTCCGCGCATCTCAACGGCGACGCAAGCGCATCGGGGTGC 1090
QY 500 ACAAGTGTCTTCTGCTCACCCTCACCACCAACATGATGGGTAAACATCAATCATCA 559
Db 1091 TGTAGACTACGGCAAGGACGACGACCCACACCGGACAGGGCTCTGGACGTTACCAGCA 1150

; TYPE: DNA
; ORGANISM: Aequoria Victoria and Human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1092)
; OTHER INFORMATION:
US-10-270-223-5

Query Match      3.3%; Score 45.8; DB 6; Length 1092;
Best Local Similarity 47.4%; Pred. No. 0.0045;
Matches 137; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 678 CAACATCAACATGTTGGATCCGTGACTTCTACATCTTCCGCAAGAGGTTGGACGGTAAGA 737
Db 291 CACCATCTTCTTCAAGGACGACGCGCACTACAAGCCGCGCGAGGTGAAGTTCGAGGG 350
QY 738 CATCAACATCTGTTCAACTCTTTCAGTACACCAAGCTGCTCAAGGACTACTTGGGTTAA 797
Db 351 CGACACCCCTGTTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGCGCAACAT 410
QY 798 CGACCTGAGATACACAGGAGTACTACATGGTCAACATCGACTACTTGNACAGATACAT 857
Db 411 COTGGGGCACAAGTGGAGTACAACACTACAACAGCCCAACAGCTGTATATCATGCGCGACAA 470
QY 858 GTACCCCAACTCCAGACAGATCGTCTTCAACACCAGACGTTAAACAACAACGACTTCAACGA 917
Db 471 GCAGAAGAAGCGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAGGACGCGAGCGT 530
QY 918 GGGTTACAAGATCATCATCAAGCTGTTTCAGAGGTAAACCAACCAAGACAC 966
Db 531 GCAGCTCCGACCACTACCAGCAGACACACCCCATCGGCGAGCGGCC 579

RESULT 14
US-10-230-437-123
; Sequence 123, Application US/10230437
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3530P1C94
; CURRENT APPLICATION NUMBER: US/10/230,437
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 10/119,480
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063549
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/069873
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
```

```
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 246
; SEQ ID NO 123
; LENGTH: 3401
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-230-437-123

Query Match          3.2%; Score 44.4; DB 6; Length 3401;
Best Local Similarity 44.3%; Pred. No. 0.019;
Matches 180; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 730 GGTAAAGGACATCAACATCTGTTCACTCTTCCAGTACACCAAGCTGTCAGAGCTAC 789
Db 983 GGGTGCACCTCGAGAGCTGTCATCAACATGAGGGCACCACGCTCATGCTCTCAAC 1042

QY 790 TGGGGTAACGACCTGAGATACACAAAGAGTACTATCATGTTCAACATCGACTACTTGAAC 849
Db 1043 AGCCTCAAGAAGATGCGAACCTGACTGAGCTGAGCTGATCCGCTGGGACCTGGAGCGC 1102

QY 850 AGATACATGTACCCCACTCCAGACAGATCGTCTTCAACACAGACGTAAACAAACGAC 909
Db 1103 ATCCCCCACTCCATCTCAGCTTCCACACCTCGAGAGATTGACCTCAAGGACAAACAC 1162

QY 910 TTCAAGAGGGTTTACAGATCATCAAGCGTATCAGAGGTAAACACCAACGACACACGAGA 969
Db 1163 CTCAAGACCATCGAGGAGATCATCAGCTTCCAGACCTGCACCGCTTCTTACTGCGGC 1342

QY 970 GTGAGAGGTGGTACATCTGTTACTTTCGACATGACTATCAACAAAGGCGCTTACAACTG 1029
Db 1223 CTGTGGTACAAACCATCGCTTACATCCCATCCAGATCGGCAACCTCACCACCTGGAG 1282

QY 1030 TTCAATGAAGAACGAGACCATGTACGCCGACAAACCACTCCACGAGGACATCTACGCCATC 1089
Db 1283 CGCCTCTACCTGAACCGCAACAGATCGAGAGATCCCCACCCAGCTTCTTACTGCGGC 1342

QY 1090 GGTCTGGTGGAGACACCAAGGACATCAACGACACATCATCTTCC 1135
Db 1343 AAGCTGGCTACCTGGACCTCAGCCCAACAAACCTGACCTTCTCTCC 1388

RESULT 15
US-10-125-923A-411
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC79
; CURRENT APPLICATION NUMBER: US/10/125,923A
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
```

```
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 411
; LENGTH: 3401
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-125-923A-411

Query Match          3.2%; Score 44.4; DB 6; Length 3401;
Best Local Similarity 44.3%; Pred. No. 0.019;
Matches 180; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 730 GGTAAAGGACATCAACATCTGTTCACTCTTCCAGTACACCAAGCTGTCAGAGCTAC 789
Db 983 GGGTGCACCTCGAGAGCTGTCATCAACATGAGGGCACCACGCTCATGCTCTCAAC 1042

QY 790 TGGGGTAACGACCTGAGATACACAAAGAGTACTATCATGTTCAACATCGACTACTTGAAC 849
Db 1043 AGCCTCAAGAAGATGCGAACCTGACTGAGCTGAGCTGATCCGCTGGGACCTGGAGCGC 1102

QY 850 AGATACATGTACCCCACTCCAGACAGATCGTCTTCAACACAGACGTAAACAAACGAC 909
Db 1103 ATCCCCCACTCCATCTCAGCTTCCACACCTCGAGAGATTGACCTCAAGGACAAACAC 1162

QY 910 TTCAAGAGGGTTTACAGATCATCAAGCGTATCAGAGGTAAACACCAACGACACACGAGA 969
Db 1163 CTCAAGACCATCGAGGAGATCATCAGCTTCCAGACCTGCACCGCTTCTTACTGCGGC 1222

QY 970 GTGAGAGGTGGTACATCTGTTACTTTCGACATGACTATCAACAAAGGCGCTTACAACTG 1029
Db 1223 CTGTGGTACAAACCATCGCTTACATCCCATCCAGATCGGCAACCTCACCACCTGGAG 1282

QY 1030 TTCAATGAAGAACGAGACCATGTACGCCGACAAACCACTCCACGAGGACATCTACGCCATC 1089
Db 1283 CGCCTCTACCTGAACCGCAACAGATCGAGAGATCCCCACCCAGCTTCTTACTGCGGC 1342

QY 1090 GGTCTGGTGGAGACACCAAGGACATCAACGACACATCATCTTCC 1135
Db 1343 AAGCTGGCTACCTGGACCTCAGCCCAACAAACCTGACCTTCTCTCC 1388
```

Search completed: November 7, 2002, 14:42:56  
Job time : 189 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:18:45 ; Search time 1949 Seconds  
(without alignments)  
11392.515 Million cell updates/sec

Title: US-09-910-186A-9

Perfect score: 1371

Sequence: 1 gaattcaagatgaccatccc.....tctccgagtaataagaattc 1371

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estnu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_man:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_pro:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query	Description
1	107.2	7.8	641	BM181884
2	107	7.8	906	BM153606
3	106.2	7.7	816	BM1535744
4	104.4	7.6	890	BM146886
5	104.4	7.6	908	AZ548467
6	104.2	7.6	905	AZ550256

c	7	100.6	7.3	976	17	BM149983
c	8	100.4	7.3	843	17	AZ551618
c	9	100.2	7.3	891	17	AZ683582
c	10	99.4	7.3	849	17	AZ546009
c	11	98.8	7.2	494	13	BM360881
c	12	98.8	7.2	681	17	CNS02EOD
c	13	98.4	7.2	931	17	BM160272
c	14	97.4	7.1	900	17	AZ549980
c	15	96.8	7.1	797	13	BM324597
c	16	95.8	7.0	698	13	BM330059
c	17	95.4	7.0	877	17	AZ531291
c	18	95.4	7.0	912	17	AZ551092
c	19	92.6	6.8	356	13	BM337970
c	20	92.6	6.8	645	9	BM337970
c	21	92.4	6.7	603	13	BM366946
c	22	92.2	6.7	500	17	BM67199
c	23	91.8	6.7	450	13	BM371297
c	24	91.4	6.7	711	13	BM588321
c	25	90.6	6.6	551	13	BM366220
c	26	90.2	6.6	317	13	BM366180
c	27	90.2	6.6	394	9	AU060224
c	28	89.8	6.5	511	13	BM004465
c	29	89.8	6.5	827	17	CNS02156
c	30	89.4	6.5	649	13	BM641907
c	31	89.2	6.5	527	10	AV962678
c	32	88.8	6.5	880	17	AZ523191
c	33	86.2	6.3	644	13	BM605179
c	34	85.6	6.2	585	13	BM328723
c	35	85.2	6.2	888	17	AZ528430
c	36	84.8	6.2	789	17	AQ935224
c	37	84.8	6.2	890	17	AZ530768
c	38	84.2	6.1	542	13	BM594734
c	39	83.8	6.1	494	13	BM641696
c	40	83.6	6.1	359	13	BM366856
c	41	83.6	6.1	503	13	BM656118
c	42	83.4	6.1	878	17	BM153470
c	43	83.4	6.1	970	17	CNS03H6V
c	44	82.6	6.0	607	13	BM348812
c	45	81.8	6.0	608	13	BM887904

#### ALIGNMENTS

BM181884 641 bp mRNA linear EST 11-DEC-2001  
fv51b11.y1 Sugano SJD adult male Danio rerio cDNA clone 5412044 5'  
similar to contains element TAK1 repetitive element ;, mRNA  
sequence.

ACCESSION BM181884.1 GI:17512842  
VERSION BM181884  
KEYWORDS EST.

SOURCE zebrafish.  
ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 641)

AUTHORS

Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy  
S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood  
K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,  
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,  
Kohn, S., Shin, T., Jackson, X., Cardenas, M., McCann, R., Waterston, R.  
and Wilson, R.

TITLE Washu Zebrafish EST Project 1998  
JOURNAL Unpublished (1998)

COMMENT

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@watson.wustl.edu

Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA sequencing by: Washington University Genome Sequencing Center Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: [info@image.llnl.gov](mailto:info@image.llnl.gov)  
Seq primer: T3 ET from Amersham  
High quality sequence stop: 527.

**FEATURES**  
**SOURCE**

```

location/Qualifiers
1. .641
/organism="Danio rerio"
/db_xref="taxon:7955"
/cclone="5412044"
/cclone.lib="Sugano STD ad
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage r
/note=vector: pME18S-FL3
site.2: Oligo1 (CACATGTG
with an oligo(dT) primer
double-stranded cDNA was
[TGTTGCCACTG], digeste
sites of the pME18S-FL3 v
CACATGTG). XhoI should b
insert. Size selection wa
<1.kb. Library construct
Sugano (University of Tok
Custom primers for sequen
CTTGTGCTTAAAGCTGCG and
CGACTGTCAGCTCGACACA.
234 37 152

```

BASE COUNT

	COACCTCCAGGTCGGACCA.		
218 a	234 c	37 g	152 t

Query Match	7.8%; Score 107.2; DB 13; Length 641;
Best Local Similarity	49.7%; Pred. No. 3.9e-13;
Matches 308; Conservative	0; Mismatches 103; Indels 9; Gaps 1;

  

QY	270	CATCGTCTCAAACTCCCATGTACGAGTGCTTTCTCCATCCTCTTCGGATCAGAATCAACA	329
DB			
Db	30	CATCATCTCCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTA	89
QY	330	GTGGGTCTCCAACTTTGCCAGGTTTAACCATCATGCGATCGGTCGAAGAACAATCCGGTTG	389
DB			
Db	90	CTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTGCTACTCAATT	149
QY	390	GTCATCGGTATCATCTCCAACTTCCCGTCTTCCACCCTGAACGACGAAGAGGATCCGA	449
DB			
Db	150	CTCCAACAACAACAACCTTCTCCAGTACATCAACTTCTCCAACAACAACAACCTTCTCAA	209
QY	450	GCAGTCCATCAACTTCTCCTACGACATCTCCAACAACGCTCCTGGTTACAACAAGTGTT	509
DB			
Db	210	CAACAACAACCTTCTCCAGTACATCAACTTCACAACA ----- ACACAACCTTCTC	260
QY	510	CTTGTGTCACCGTCCACAACAACATGATGGGTACATGAAGATATACATCAACGGTAAGCT	569
DB			
Db	261	CAGCTACATCAACTTTCACAACAACAACCTTCTCCAGTACAACTTCTCCAGCAACAACAT	320
QY	570	GATCGACACCATCAAGGTCAGGAGTTGACGGGTATCACTTCTCCAAGACCATCACCTT	629
DB			
Db	321	CAACTTCTCCAGTACATCAACTTCTCCAACAACAACAACCTTCTCCAGTACATCAACTT	380
QY	630	CGAGATCAACAAGATCCCAAGACACGGTCTGATCACTCCGATCCGACCAACATCAACAT	689
DB			
Db	381	CTCCAACAACAACACTTCTCCAGTACATCAACTTCTCCAACAACAACAACCTTCTCCAG	440
QY	690	GTGGATCCGTGACTTCTACATCTTTCGCCAAGGAGTTGGAGGTPAAGGACATCAAGATCCT	749
DB			
Db	441	CTACATCAACTTTCACAACAACAACAACCTTCTCCAGTACATCAACTTCAACAACAACA	500
QY	750	GTTCAACTCCTTGCAGTACACCAACGCTCGTAGGCTACTGGGGTAAGGACCTGAGATA	809
DB			
Db	501	CTTCTCCAGCTACAACCTTCTCCAGCAACAACATCAACTTCTCCAGTACATCAACTTCAA	560
QY	810	CAACAGGAGTACTACATGGTCAACATCGACTACTTGAACAGATACATGATGCCAACCTC	869

Db	561	CAACACACAACTTCTCCAGCTCATCACTTCAACACACAACTTCCAGCTA	620
Qy	870	CAGACAGATCGTCTTCAACA	889
Db	621	CAACTTCTCCAGCAACAACA	640

  

RESULT 2	BH153606	906 bp	DNA	linear	GSS 24-SEP-2001
LOCUS	ENTTS583TF	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica	
DEFINITION	genomic, DNA sequence.				
ACCESSION	BH153606				
VERSION	BH153606.1	GI:15725323			
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica.				
ORGANISM	Entamoeba histolytica				
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.				
AUTHORS	Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.				
TITLE	Determination of clone end sequences from Entamoeba histolytica				
JOURNAL	HMI:IMSS sheared DNA library (2001)				
COMMENT	Unpublished (2001)				
	Contact: Brendan J Loftus				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0208				
	Fax: 301 838 3543				
	Email: bjl@loftus81.org				
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared				
	DNA library				
	Seq primer: M13-Forward				
	Class: shotgun				
	High quality sequence start: 15				
	High quality sequence stop: 733.				

## FEATURES

1. .906  
source  
/organism="Entamoeba histolytica"  
/strain="HM1:IMS"  
/db\_xref="taxon:5759"  
/clone.lib="Entamoeba histolytica Sheared DNA"  
/note=|vector: pEOS1; Site\_1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."  
149 a 293 c 66 g 398 t  
BASE COUNT  
ORIGINAL

[illegible]

Query Match	7.8%	Score 107;	DB 17;	Length 906;
Best Local Similarity	46.4%;	Pred. No. 5e-13;		
Matches 350;	Conservative 0;	Mismatches 405;	Indels 0;	Gaps 0;
QY	247	GTGTCGGTCACCCAGCAAGCAGCAAGATCGTCTACAACTCCCATGTACGAGTCCTCTCCCATC	306	
Db	41	GTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATC	100	
QY	307	TCCTTTGGATCAGAACTCAACAAGTGGGTCTCCAACTTGCAGGTTACACCACTATCGAC	366	
Db	101	GTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATC	160	
QY	367	TCGGTCAAGAAACAATCCGGTGGTCCATCGGTATCATCTCCAACTTCGTGGTCTTCACC	426	

```

Db 161 GTCTTCTTCATCGTCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 220
QY 427 CTGAAGCAGAACGAGGAGCCGAGAGTCCATCAACTTCTCTCAGAGATCTCCCAACAAAC 486
Db 221 GTGCTTCTTCATCGTCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 280
QY 487 GCWCTGGTTACAGAAAGTGGTCTTCGTCACCGGTACAGCAACATGATGGTAAATG 546
Db 281 GTCTTCTTCATCGATCATCTTCATCACTTCTTCATCGCTTCTTCATCGCTTCTTCATC 340
QY 547 AAGATCTACATCAACCGGTAAAGTGTATCGAGCACCACATCAAGGTGAGCGGTATC 606
Db 341 ATGCTCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 400
QY 607 AACTTCTCCAGACATCACTCTTCGAGATCAACAGATCCCAAGACACCGGTCTGATCAC 666
Db 401 ATGCTCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 460
QY 667 TCGAGTCCGACACATCAACATGTGAGTGGTCTTCATCGCTTCTTCATCGCTTCTTCATC 726
Db 461 ATGCTCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 520
QY 727 GAGGTAAGGACATCACTCTTCGTTTCAACTCTTCGAGTACCAACAGTCTGTCAGGAC 786
Db 521 ATGCTCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 580
QY 787 TAGTGGGTAACGACCTGAGATACACAGAGGATCTACATCGTCAACATCGCTTCTTCATC 846
Db 581 TTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 640
QY 847 AACAGATACATGATCGGACCTGAGATACAGAGGATCTACATCGTCAACATCGCTTCTTCATC 906
Db 641 GTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 700
QY 907 GACTTCAACGAGGTTACAGATCATCATCAACGCTATCAGAGGTAACACAGCAGACACC 966
Db 701 TTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 760
QY 967 AGAGTCAGAGGTTGACATCTCTGATCTCTGACAT 1001
Db 761 ATCATCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATC 795

RESULT 3
AZ535744
LOCUS
DEFINITION
ENTCQ25TR Entamoeba histolytica 816 bp DNA linear GSS 03-NOV-2000
genomic, DNA sequence.
ACCESSION
AZ535744
VERSION
AZ535744.1 GI:11092691
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica.
ORGANISM
Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
1 (bases 1 to 816)
Loftus B., Van Aken S. and Fraser C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 36
High quality sequence stop: 816.
Location/Qualifiers

```

```

source
1..816
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica sheared DNA"
/notes="Vector: pGEM1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT 125 a 265 c 54 g 372 t
ORIGIN
Query Match 7.7%; Score 106.2; DB 17; Length 816;
Best Local Similarity 46.9%; Pred. No. 7.2e-13;
Matches 330; Conservative 0; Mismatches 373; Indels 0; Gaps 0;
QY 15 CATCCCATCAACATCTTCTCTACACCAACATCCCTGTGTGAAGGACATCATCAACGA 74
Db 105 CATCGTCGTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCAT 164
QY 75 GTACTTCAACATCAACGATCTCCAGATCTGCTCCCTGAGAACCGTAGAACACCTT 134
Db 165 CATCATCATCATCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCTTCTTCATCGCT 224
QY 135 GGTGACACCTCCGGTTACAAGCCGAGGTCTCCGAGGAGGTGACGCCAGCTGAAGCC 194
Db 225 CTTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTT 284
QY 195 AATCTTCCCATTCGATCTCAAGTGGTTCCTCCGGTGAAGGACAGGATAGGCACTGT 254
Db 285 CTTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTT 344
QY 255 CAGCCAGAGAGAACATCGTCTACAACTCCATGTACAGAGTCTCTCTCTCTCTCTCTCT 314
Db 345 CTTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTT 404
QY 315 GATCAGAAATCAAAAGTGGTCTCCAACTTCCAGGTATACACCATCATCGCTCGTCAA 374
Db 405 CTTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTT 464
QY 375 GAACAACTCCGGTTGCTCCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTT 434
Db 465 CATCGCTTCTTCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTTCTTCATCGCTT 524
QY 435 GAACGAGGATCCGAGCAGTCCATCAACTTCTCCAGCATCTCCAAAGAGGTCCTCGG 494
Db 525 CATCGCTTCTTCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCG 584
QY 495 TTACAAAGTGGTCTTCGTCACCGTCCAGACATGATGGGTAAACATGAAGATCTA 554
Db 585 CATCTTCTTCATCATCGCTTCTTCATCATCGCTTCTTCATCATCGCTTCTTCATCAT 644
QY 555 CATCAAGGTAAGTGTGACACCATCAAGGTGAAGGATGACCGGTATCAACTTCTC 614
Db 645 CATCATCATCTTCTTCATCATCATCTTCTTCATCATCATCTTCTTCATCATCATCGT 704
QY 615 CAGACCATCACTTCCGATCAACAGATCCAGACACCGGTCTGATCACTCCGACCTC 674
Db 705 CTTTCATCATCATCTTCTTCATCATCATCTTCTTCATCATCATCTTCTTCATCATCAT 764
QY 675 CGACAAATCAACATGTGGATCCGTCGCTGACTTCTACATCTTCGCTCGCC 717
Db 765 CTGCATCATCATCTTCTTCATCATCATCTTCTTCATCATCATCTTCTTCATCATCATCG 807

```

```

RESULT 4
BH146886/c
LOCUS      BH146886      890 bp      DNA      linear      GSS 27-AUG-2001
DEFINITION  ENTPE487F Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION  BH146886
VERSION    BH146886.1 GI:15302963
KEYWORDS   Entamoeba histolytica.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica.
REFERENCE  1 (bases 1 to 890)
AUTHORS    Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library (2001)
JOURNAL    Unpublished (2001)
COMMENT    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 880.
FEATURES   source
            Location/Qualifiers
                1..890
                /organism="Entamoeba histolytica"
                /strain="HMI:IMSS"
                /db_xref="taxon:5759"
                /clone_lib="Entamoeba histolytica Sheared DNA"
                /note="Vector: PHOS1; Site_1: Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999)."
```

```

BASE COUNT 404 a 55 c 285 g 136 t
ORIGIN
Query Match      7.6%; Score 104.4; DB 17; Length 890;
Best Local Similarity 45.6%; Pred. No. 1.8e-12;
Matches 369; Conservative 0; Mismatches 441; Indels 0; Gaps 0;
QY 223 TCCGCGGTGAGGACAGGTAGGTCATCGTCACCCAGACAGACATCGTACAC 282
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841 TTCTCTGCGTCTTAATCAATTCATCATCGTCTTCTTAATCAATTCATCGTC 782
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 283 TCCATGTACGAGTCCCTCTCCATCTCTCTTGGATCAGATCAACAAGTGGTCTCCAA 342
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 TTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTC 722
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 TTGCCAGGTACACCATCAGTCAGTCGTCGACAGAACTCCCGTTGGTCCATCGTATC 402
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 TTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTC 662
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 403 ATCTCCAACTTCCTGGTCTTCAACCTCGACGAGACGACGCCAGCATCCATCAAC 462
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 TTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTC 602
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 463 TTCTCCACGACATCTCCACAGCGTCCGTTTACACAAGTGGTCTTCGTCCCGTC 522
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 TTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCATCA 542
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 523 ACCAAACAACATGATGGTGAACATGAAGTACTACATCAACGGTAAAGCTGATCGACACCA 582
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 TTCTTCATCGTCTTCTTCATCATCATCATCTTTCATCATCTTCTTCATCATCGTC 482
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 583 AAGTCAAGGAGTTGACCGGTATCAACTTCTCCAGAGACCATCACCTTCGAGATCAACAAG 642
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 481 TTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTC 422
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 643 ATCCAGACACCGGTCTGATCACTCCGACTCCGACATCAACATCAAGATGTGATCGGTGAC 702
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 TTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTC 362
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 703 TTCTACATCTTCCGCAAGGAGTTGGAGCTGAAGGATCAACATCTGTTCAACTCCCTTG 762
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 TTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTC 302
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 763 CAGTACACCAACGTCTGTCAGGAGTACTGCGGTAAAGCTGAGATCAACAAGGAGTAC 822
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 TTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCAATTC 242
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 823 TACATGTCACATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCAATTC 882
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCGTC 182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 883 TTCAACACCAAGAGTAAACACAGACGACTTCAAGAGGGTTACAAGATCATCATCAAGCGT 942
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 TTCTTCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTC 122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 943 ATCAGAGGTAAACACCAAGACACGAGTCAAGAGTGGTGACATCTGTTACGTCAGATG 1002
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ATCATCATCTTCTTCATCATCATCGTCTTCTTCATCATCATCGTCTTCTTCATCATCATC 62
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1003 ACTATCAACACCAAGGCCCTACACACCTGTC 1032
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TTCTTCATCATCATCTTCTTCATCATCATC 32
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 5
AZ548467/c
LOCUS      AZ548467      908 bp      DNA      linear      GSS 14-NOV-2000
DEFINITION  ENTER30RF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION  AZ548467
VERSION    AZ548467.1 GI:11172102
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica.
ORGANISM   Entamoeba histolytica.
REFERENCE  1 (bases 1 to 908)
AUTHORS    Loftus,B., Van Aken,S. and Fraser,C.
TITLE      Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library
            Unpublished (2000)
JOURNAL    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 828.
FEATURES   source
            Location/Qualifiers
                1..908
                /organism="Entamoeba histolytica"
                /strain="HMI:IMSS"
```









Query Match	7.38;	Score 100.2;	DB 17;	Length 891;
Best Local Similarity	46.3%;	Pred. No. 1.5e-11;		
Matches 367;	Conservative	0;	Mismatches 423;	Indels 3; Gaps 1;

  

QY	289	TACGAGTCCTTCTCCATCCTCCTTGATCAGATCAACAAGTGGGTCTCCAACATTGCCA	348	
DB	796	TAATCGTAGTTCTTCATCGTCTTTATCCATCGCTCTTCTTCAATCGTCACTTCATGTGCT	737	
QY	349	GATTACACCATCATCGACTCCGTTCAAAGAACAATCCGGTTGGTCCATCGGTATCATCTCC	408	
DB	736	TCCTTCATCGTCTTCTTCATCGTCTTCTCCATCGTCTGCTCTCATCGTCTGCTTCATCGTC	677	
QY	409	AATCTCTCTGGTCTTACACCTGAAGCAAGAAGAGACTCCGAGCAGTCCATCAACTTCCC	468	
DB	676	GTCTTCATCGTCTTTCATCATCTTCTTCATCGTCTTCTTTCATCATCATCATCTTCATC	617	
QY	469	TACGACATCTCAACAACGCCTCTGTTAAACAAGTGGTCTTCTGTCACCGTCAACCAC	528	
DB	616	AUCTTCACTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC	557	
QY	529	AACATGATGGGTAAACATGAAGATCTACATCAACGGTAAAGTATGACACCAATCAAGTCC	588	
DB	556	ATCATCGTCTTCTTTCATCATCGTCTTCTTTCATCATCGTCTTCTTTCATCATCGTCTTCTTC	497	
QY	589	AAGGAGTTTGACCGGTATCAACTTCTCAAGACATCATCCTTCGAGATCAACAAGATCCCA	648	
DB	496	ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC	437	
QY	649	GACACGGTCTTGATCACCTCCGACTCCGACAAACATCAACATGATGGATCGGTGACTCTTA	707	
DB	436	ATCATCGTCTTCTTTCATCATCGTCTTCTTTCATCATCGTCTTCTTTCATCATCGTCTTCTTAA	377	
QY	708	-CATCTTTCGCAAGAGTTGGAGGGTAAAGACATCAACATCTGTTTCAACTCCCTTGCAG	765	
DB	376	TTCAATTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCGTCTTC	317	
QY	766	TACACAACGTGCTCAAGGATCTACTGGGGTAAACACCTGATGATAACAAGGAGTACTAC	825	
DB	316	TTCATCGTCTTCTTTCATCGTCTTCTTCATCGTCTTCTTTCATCGTCTTCTTTCATCGTCTTC	257	
QY	826	ATGGTCAACATCGATCTACTTGAACAGATACATGTATCGCCAACTPCCAGACAGATCGTCTTC	885	
DB	256	TTTCATCGTCAATCATCTTCTTTCATCATCATCGTCTTCTTTCATCATCATCGTCTTCATCTTC	197	
QY	886	AACACAGACGTTAAACAACAAGCACTTCAAGGAGTTTACAGAGTATCATCATCAAGGATTC	945	
DB	196	ATCATCATCTTCTTTCATCATCATCTTCTTTCATCATCATCTTCTTTCATCATCATCTTCTTC	137	
QY	946	AGAGGTAAACCAACGACACCAAGTACAGAGGTGGTGACATCTCTGCTACTTGCACATGACT	1005	
DB	136	ATCATCATCGTCTTTCATCATCATCTTCTTTCATCATCATCTTCTTTCATCATCATCTTCTTAA	77	
QY	1006	ATCAACAACAAGGCTTCAACCTGTTCATGAAGACGAGACCACTGTACGGCAGACACCAC	1065	
DB	76	TTCAATTCATCATCTTTCATCATCAATCATCTTCATCATCATCTTCTTTCATCATCTTCTTC	17	
QY	1066	TCCACCGAGGACA	1078	
DB	16	ACTTCCACGACACA	4	

RESULT 10  
AZ546009  
LOCUS  
DEFINITION  
AZ546009  
ENTPW53TF Entamoeba histolytica Sheared DNA  
genomic, DNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

849 bp DNA linear GSS 14-NOV-2000  
Entamoeba histolytica  
Entamoeba histolytica  
GSS.  
Entamoeba histolytica.  
Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.  
1 (bases 1 to 849)

AUTHORS	Loftus, B., Van Aker, S. and Fraser, C.
TITLE	Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Brendan J. Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@tigr.org Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library Seq primer: M13-Forward Class: shotgun High quality sequence start: 26 High quality sequence stop: 796. Location/Qualifiers 1. .849
FEATURES	source

1. .849  
/organism="Entamoeba histolytica"  
/strain="HMI:IMS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/note="vector: PHO1; Site\_1: Bst I; Constructed at the  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell. Oxford University Press, 1999)."

```

BASE COUNT      199 a      236 c      43 g      371 t
ORIGIN

Query Match      7.3%; Score 99.4; DB 17; Length 849;
Best Local Similarity 47.7%; Pred. No. 2.2e-11;
Matches 326; Conservative 0; Mismatches 351; Indels 6; Gaps 1;

QY  270 CATCGCTACAACTCCATGCTAGGAGTCCTCTCCATCTCTTGGATCAGATCAACAA 329
DB  49 CGTCACTTCAATCAACACTCTTCGTCATCTTCATCATCTTCATCATCAT 108
QY  330 GTGGGCTCCAACTTSCGAGTTACACATCATCGATCCGGTCAAGAACAACTCCGGTTG 389
DB  109 CTTCTCTTCACTCTTCATCATCATCTTCTCTTCTTCTTCTTCTTCATCATCTTCATCTCTCT 168
QY  390 GTCCATCGGTATCANTCCCACTTCTCGTCTTCCACCTGAAGCAGAAAGGAGCTCGGA 449
DB  169 CTTCACTTCACTCTTCATCTTCACTCTTCACTCTTAAATTCATCATCATCTTCGTCTCTT 228
QY  450 GCAGTCCATCAACTTCTCCTAGGACATCTCCAAACGCTCCGGTTACAAACAGTGTTT 509
DB  229 CGTCTTCATCATCATCTTTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 288
QY  510 CTTTCGTCCGGTCCACCACAAACATGA-----TGGGTAAACATGAAGATCFACATCAACGG 563
DB  289 CTTCTTCATCTTCACTCATCATCATCATCTTCATCATCTTCACTTCATCATCTTCATCATCTT 348
QY  564 TAAAGTGTATCGACACCATCAAGTCAAGGAGTTGACCGGTATCAACTTCTCCAAAGACCAT 623
DB  349 CATCATCTTTCATCATCTTTCATCATCTTTCATCATCATCATCATCTTCTTCTTCATCTTCTTCT 408
QY  624 CACCTTCGAGATCAACAGATCCAGACACCGGTCGTGATCACTCCGAGCTCCGACAAACAT 683
DB  409 CATCTCTTCTCACTCTTCTTAAATTCAAATTCATCATCATCATCTTCTTCTTCTTCTTCTTCTTCT 468
QY  684 CAACATGTGGATCCGCTGACTTCTACATCTTCGCCAAGAGGTTGGAGCGGTAAGACATCAA 743
DB  409 CATCTCTTCTCACTCTTCTTAAATTCAAATTCATCATCATCATCTTCTTCTTCTTCTTCTTCTTCT 468

```

Db 459 CATCATCATCTCTTCTTCATCTTCATCATCTCTTCATCATCATCTCTTCTTCATCATCAT 528  
Qy 744 CATCTCTTCAACTCTCTTGGAGTACACCAAGCTGCTGCAAGGACTACTGGGGTAAGCACT 803  
Db 529 CATCTTCATCTCTTCTTCATCTCTTCACTCTTCACTCAATCAATCATCATCTCTTCTTCTTCT 588  
Qy 804 GAGATACACAGAGAGTACTACATCTGCTCAACATCGACTACTTGAACAGATACATGTACGC 863  
Db 589 CATCATCATCATCTCTTCTTCTTCATCTCTTCTGCTCTTCATCATCATCATCATCATCTCTTCT 648  
Qy 864 CAATCCAGACAGATCTCTTCTCAACACAGACGCTTAACAACAGACTTCAACGAGGGTTA 923  
Db 649 CTCTCAATCAATTCATCATCATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 708  
Qy 924 CAAGATCATCATCAAGCGTATCA 946  
Db 709 CATGTTCCATTTCTACATTTTCA 731

RESULT 11  
LOCUS BJ360881  
DEFINITION BJ360881 Dictyostelium discoidium cDNA library, CF Dictyostelium  
ACCESSION BJ360881.1 GI:19260476  
VERSION BJ360881  
KEYWORDS EST  
SOURCE Dictyostelium discoidium  
ORGANISM Dictyostelium discoidium  
REFERENCE Eukaryote; Mycetozoa; Dictyostelidia; Dictyostelium.  
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.  
TITLE Full length cDNA of Dictyostelium discoidium at the culmination stage  
JOURNAL Unpublished (2002)  
COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1. .494  
/organism="Dictyostelium discoidium"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="ddc8k12"  
/clone\_lib="Dictyostelium discoidium cDNA library, CF"  
/sex="mat A"  
/dev\_stage="Culmination stage"  
BASE COUNT 279 a 129 c 34 g 48 t 4 others  
ORIGIN

Query Match 7.2%; Score 98.8; DB 13; Length 494;  
Best Local Similarity 52.2%; Pred. No. 2.4e-11;  
Matches 214; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

Qy 680 ACATCAACATGTGGATCGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 739  
Db 80 ACATGAAGATATAGATTGGTTTAAATTTCTTGTATAGTTTACATTCATCTTGTATACGA 139  
Qy 740 TCAACATCTCTTCACTCACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 799  
Db 140 ACAAATTCAC 199  
Qy 800 ACTGAGATAC 859  
Db 200 ACA 259  
Qy 860 AGCCAACTCCAGACAGATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 919  
Db 260 GCAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 319

Qy 920 GTTACAGATCATCATCAAGCGTATCAGAGGTAAACACCAAGCAGACACCAAGTCAAGGTG 979  
Db 320 GCAACAAGACAGAC 379  
Qy 980 GTGACATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1039  
Db 380 AC 439  
Qy 1040 ACGAGACCATGTAGCGGAC 1089  
Db 440 ACA 489

RESULT 12  
CNS02EOD  
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone  
DEFINITION 262H14 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
ACCESSION AL193990  
VERSION AL193990.1 GI:7832096  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetraodon nigroviridis  
ORGANISM Tetraodon nigroviridis  
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
AUTHORS 1 (bases 1 to 681)  
Roest-Crolius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,  
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,  
Saurin, W. and Weissenbach, J.  
TITLE Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 681)  
Roest-Crolius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,  
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and  
Weissenbach, J.  
TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
JOURNAL Unpublished  
AUTHORS 3 (bases 1 to 681)  
Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (12-APR-2000)  
COMMENT This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
Location/Qualifiers

FEATURES  
source  
1. .681  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="262H14"  
/clone\_lib="G"  
/notes="Genoscope sequence ID : C0AG262DD07LPI-end : T7"  
BASE COUNT 74 a 105 c 170 g 282 t 50 others  
ORIGIN

Query Match 7.2%; Score 98.8; DB 17; Length 681;  
Best Local Similarity 48.8%; Pred. No. 2.7e-11;  
Matches 197; Conservative 20; Mismatches 187; Indels 0; Gaps 0;

Qy 813 CAAGAGTACTACATGTGTCACATCGACTACTTGAACAGATACATGTACGCCAATCCAG 872  
Db 406 CDAC 347  
Qy 873 ACAGATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 932  
Db 346 CAACGAC 287

QY	933	CATCAAGCGTATCAGAGGTACACCACGACGACCCAGAGTGTCAGAGTGGTGACATCTTGTA	992
Db	286	CAGCGACGACGACGACGACGACAAACAAAACRGCACACRACGCGGACRACAACAACAA	227
QY	993	CTTCGACATGACTATCAAACAACGAGGCGCTACAACCTGTTTCATGAAGAAGACGACCATGTA	1052
Db	226	CAACAACAACAAAAAACHAACACRGCAACACNAGCGCNACACACACNACACACRACAAAA	167
QY	1053	CGCGGACAACCACTCCACGAGGAGCATCTACGCCATCGGCTTCGGTGTGAGCAGACCAAGGA	1112
Db	166	CRACAACAACACRACGCGACAAACAACACRACAACAACAAAAACAACAACAACAACAACAA	107
QY	1113	CATCAACGACACATCATCTTCCAGATCCAGCGCAATGAACAACACTTACTACTACGGTTC	1172
Db	106	AAACAACRACAACRACRCRCRCAACACACACAAAAACACACACACGCGCACACAASMA	47
QY	1173	CCAGATCTTCAAGTCCAACTTCAACGGTGAGACATCTCCGGTA	1216
Db	46	CAACAACRACAACRACACGACACACGACACACACACACACACACCTTKA	3

RESULT 13  
BH160272/c

LOCUS BH160272 linear GSS 24-SEP-2001

DEFINITION ENTQV49TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.

ACCESSION BH160272 GI:15733710

VERSION BH160272.1

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 931)

AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library (2001)

JOURNAL Unpublished (2001)

COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0308  
Fax: 301 838 3543  
Email: bjloftus@igr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

	Best Local Similarity	47.08;	Pred. No. 3.8e-11;			
	Matches	303;	Conservative	0;	Mismatches	341;
				Indels		Gaps
QY	249	CATCGTACCCAGAACGAGACATCGTGTACAACTCCATCGAGTCTCTTCATCTC	308			
Db	662	CGTGTCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTT	603			
QY	309	CTTCTGGATCAGATCAACAGTGGGTCCCACTTGCAGGTTACACATCATCGACTC	368			
Db	602	CTTCATCATCTTCATCTTCTTCATCATCTTCATCATCTTCATCATCTTCATCAT	543			
QY	369	CGTCAAGAACAACTCGGTTGGTCCATCGGTATCATCTCCAACTTCTCTGCTTCA	428			
Db	542	CATCTTCTTCTTTCATCATCTTCTTCTTCATCATCTTCTTTCATCATCTTCTT	483			
QY	429	GHAGCAGAAGGAGCTCCGAGAGTCCATCACTTCTCTAGACATCTCGAACAGC	488			
Db	482	CTTCTTAATCATATTCTGCTGCTTCATCATCATCTTCTTCTTCATCATCTTCTT	423			
QY	489	TCCTGGTTACAAAGTGGTTCCTGGTCAACGTCACCAACATCATGATGGTAACTG	548			
Db	422	CATCTTCTTCTTTCATCATCTTCTTCTTCATCATCATCTTCTTCTTCATCATCT	363			
QY	549	GATCTACATCAACGGTAACTGATCGACACCATCAAGGTTCACGGGTATCAA	608			
Db	362	CTTCTTCTTCATCATCTTCATCTTCTTCTTCATCATCTTCATCATCTTCTT	303			
QY	609	CTTCTCCAGACATCACTTCGAGATCAACAAGATCCACAGACCGGTCTGATCACC	668			
Db	302	CTTCTCATCATCATCTTCTTCTTCTTCATCATCATCTTCTTCTTCTTCTTCTT	243			
QY	669	CGACTCCGACACATCAACATGTGGATCGGTGACTTCTACATCTTCGCCAAGGTTGA	728			
Db	242	CATCTGCTTCATCATCTTCATCATCATCTTCTTCTTCTTCATCATCATCTTCT	183			
QY	729	CGGTAAAGACATCAACATCTCTGTTTCACTCTCTGAGTACCAACAGTCTGTCAG	788			
Db	182	CTTCATCATCTTCTTCTTCTTCTTCTTCATCATCATCTTCTTCATCATCTTCTT	123			
QY	789	CTGGGTAAAGACCTTGAGATCAACAAGGAGTACTACATGGTCAACATCGACTTGA	848			
Db	122	CATCATCATCTTCTTCTTCATCATCATCATCTTCTTCTTCTTCTTCTTCTTCT	63			
QY	849	CAGATACATGTAAGCCAACTCCAGACAGATCGTCTTCAACACCA	892			
Db	62	CGTCTTCATCATCTTTCATCATCATCATCTTCTTCTTCTTCTTCTTCTTCTTCT	19			

FEATURES	source
Class: shotgun	
High quality sequence start: 8	
High quality sequence stop: 677.	
Location/Qualifiers	
1. 931	
/organism="Entamoeba histolytica"	
/strain="HM1:IMSS"	
/db_xref="taxon:5759"	
/clone_lib="Entamoeba histolytica Sheared DNA"	
/note="Vector: pHOS1; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.	
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barell, Oxford University Press, 1999)."	
402 a 51 C 279 g 199 t	
BASE COUNT	
ORIGIN	
7.2%; Score 98.4; DB 17; Length 931;	
* Query Match	

Seq primer: M13-Forward

Class: Shotgun

High quality sequence start: 20

High quality sequence stop: 890.

## FEATURES

source

1. 900

/organism="Entamoeba histolytica"

/clone="FMI:IMSS"

/db\_xref="taxon:5759"

/note="Vector: pHOS1; Site 1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

376 a 63 c 235 g 226 t

BASE COUNT

ORIGIN

Query Match 7.1%; Score 97.4; DB 17; Length 900;

Best Local Similarity 49.0%; Pred. No. 6.2e-11;

Matches 260; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

QY 251 TGTGATCAGATCAACAAGTGGTCTCCACATGCGAGGTACACCATCATCGCTCG 370

DB 860 TCTTCATCATCGTTCATCATCGTTCATCATCGTTCATCATCGTTCATCATCG 801

QY 311 TCTGATCAGATCAACAAGTGGTCTCCACATGCGAGGTACACCATCATCGCTCG 370

DB 800 TCTTCATCATCGTTCATCATCGTTCATCATCGTTCATCATCGTTCATCATCG 741

QY 371 TCAAGAACATCGGTTGGTCCATCGGTATCATCTCCAACTTCCTGGTTCACCCCTGA 430

DB 740 TCTTCATCATCGTTCATCATCGTTCATCATCGTTCATCATCGTTCATCATCG 681

QY 431 AGCAAGAGAGACTCCGAGTCCATCATCTCTCTAGGATCTCCAGCATCTCCAAAGGCTC 490

DB 680 TCTTCATCATCGTTCATCATCGTTCATCATCGTTCATCATCGTTCATCATCG 621

QY 491 CTGGTTACAAAGTGGTCTCTCGTCCGTCACCAACATGATGGGTAACATGAAGA 550

DB 620 TCTTCATCATCTCTCTCATCATCTCTCATCTCTCTCATCTCTCTCATCTCT 561

QY 551 TCTAGATCAAGGTAAGTGTATGACACCATCAAGGTCAAGGTGACCGGTATCAACT 610

DB 560 TCAATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCA 501

QY 611 TCTCAAGCATCATCACTTCGAGATCAACAAGATCCGAGACACCGGTCATCACTCCG 670

DB 500 TCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCTCTCTTCA 441

QY 671 ACTCGGCAACATCAACATGTGATGGATCGGTACTTCTACATCTTCGCGAAGGATTTGAGC 730

DB 440 TCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCATCTTCATCT 381

QY 731 GTAAGGATCAACATCTGTTCACCTTCCTGAGTACACCAAGCTGCTCA 781

DB 380 TCAATCGTTCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 330

RESULT 15

BJ324597

LOCUS

DEFINITION BJ324597 Dictyostelium discoideum cDNA library, AF Dictyostelium

discoideum cDNA clone dda7g04 5', mRNA sequence.

ACCESSION

BJ324597

VERSION BJ324597.1 GI:19154727

KEYWORDS EST.

SOURCE Dictyostelium discoideum.

ORGANISM Dictyostelium discoideum

REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.

AUTHORS 1 (bases 1 to 797)

TITLE Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.

Full length cDNA of Dictyostelium discoideum at the aggregation stage

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshinigenes.nig.ac.jp.

Location/Qualifiers

1. 797

/organism="Dictyostelium discoideum"

/strain="AX4"

/db\_xref="taxon:44689"

/clone="dda7g04"

/clone\_lib="Dictyostelium discoideum cDNA library, AF"

/sex="mat A"

/dev\_stage="Aggregation stage"

BASE COUNT 437 a 240 c 26 g 92 t

ORIGIN

Query Match 7.1%; Score 96.8; DB 13; Length 797;

Best Local Similarity 47.0%; Pred. No. 7.9e-11;

Matches 333; Conservative 0; Mismatches 372; Indels 3; Gaps 1;

QY 422 TCACCTGAACACAGGAGGACTCGAGCAGTCCATCACTTCTCTCTCTCTCTCTCTCTCT 481

DB 32 TCAAAATCAAGCATCATATATCCAAACACACAAATTCATCATCAACACATTCGATCA 91

QY 482 ACAAGCTCTCTGTTACAAAGTGGTTCCTGTCACCGTCCACCAACACATGATGGGTA 541

DB 92 ACAACAAATGCAACAAACAAATGCAACAAACAAATGCAACAAACAAATGCAACAA 151

QY 542 ACATCAAGATCTAGATCAACGGTAAAGTGCATGACCAACATCAAGGTCAAGGTTCACCG 601

DB 152 ACACAAATGCAACAAATGCAACAAACAAATGCAACAAACAAATGCAACAAATGCA 211

QY 602 GTATCAACTTCTCAAGACCATCATACCTTCGAGATCAACAAAGTCCGACACACGGTCTGA 561

DB 212 GCAACAAACAAATCCAAATCCAAATCAACAAACAAACAAATCCAAATCAACAAACA 271

QY 652 TCACCTCGACTCCGACACATCAACATGTGGATCGGTGACTTCTCTCTCTCTCTCTCT 721

DB 272 ACAACAAATCCAAATCAACAAACAAATCCAAATCAACAAACAAATCCAAATCAACAA 331

QY 722 AGTTGGACGGTAAGGACATCAACATCTCTGTTCACCTCC---TTGCACTACACCAACGTCG 778

DB 332 AATTCACAAACCAACAAACAAATCCAAATCAACAAACAAATCCAAATCAACAAACAA 391

QY 779 TCAGGACTACTGGGTTACGACCTGAGATACACACAGGAGTACTATCGTCAACATCG 838

DB 352 TCCAAACCAACAAACAAATCCAAATCCAAATCAACAAACAAATCCAAATCAACAAACA 451

QY 839 ACTACTTGAACAGATATAGTATACCGCAACTCCAGACAGATCGTCTTCAACACACGACGTA 898

DB 452 TCCAAACCAACAAACAAATCCAAATCCAAATCAACAAACAAATCCAAATCAACAAATCC 511

QY 899 ACAACACGATCTCAAGAGGTTACAGATCATCATCAAGCGGTATCAGAGGTAAACACCA 958

DB 512 AAACCTACACACATCCAAATGCAACAAACATCCAAATGCAACAAACATCCAAATGCAACA 571

QY 959 AGGACACACAGATCAGAGGTGGTGCATCTCTGTACTTCTGACATGACTTCAACAAACG 1018

DB 572 TCCAAATGATCAACACATCCAAATGCAATCCATCCATCCATCCATCCATCCATCCATCC 631

QY	1019	CCTACAACTGTTTCATGAAGAACGAGACATGTACGCGGACACCACTCTCACCGAGACA	1078
Db	632	TGCGCCACCTCTAAATTCATTACAATGCACCAACACAAATCCAAACCAACAACA	691
QY	1079	TCTACGCCATCTGGTCTGGTGTAGCAGACCAAGGACATCAACGACA	1126
Db	692	ACAACAACCAACAAAATTCAAATCAACAACAACCAACCCAAAATCCA	739

Search completed: November 7, 2002, 13:45:20  
Job time : 1959 secs